

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**HIV-2 nucleotide sequences**

Patent Number: EP0750041

Publication

date: 1996-12-27

Inventor(s): ALIZON MARC (FR); SONIGO PIERRE (FR); GUETARD DENISE (FR); MONTAGNIER LUC (FR); TIOLLAIS PIERRE (FR); CHAKRABARTI LISA (FR); GUYADER MIREILLE (FR); DESROSIERS RONALD (US); CLAVEL FRANCOISE (US)

Applicant(s):: CENTRE NAT RECH SCIENT (FR); PASTEUR INSTITUT (FR)

Requested

Patent: ☐ EP0750041, A3

Application

Number: EP19960108720 19880115

Priority Number (s): EP19880400084 19880115; FR19870001739 19870211; FR19870005398 19870415; US19870003764 19870116

IPC

Classification: C12N15/49 ; C07K14/16 ; C12Q1/68

EC

Classification: C07K14/16B, C07K14/16D, C07K14/16, C12Q1/70B2B

Equivalents:

---

**Abstract**

---

New peptides (I) have immunological properties in common with those of the peptide skeleton of the envelope protein of HIV-2 and also have a peptide structure in common with that of SIV (simian immunodeficiency virus)-7 glycoprotein. Also new are (i) a sequence of 9600 nucleotides corresponding to the SIV genome (reproduced in the specification, together with derived amino acid sequence of viral proteins gag, pol, env, Q, X, R, tat, art and F genes) and its fragments; (2) recombinant DNA contg. all or part of the CDNA from this sequence inserted into a vector, and (3) antigenic and immunogenic conjugates contg. (I).

---

Data supplied from the **esp@cenet** database - I2



(11) **EP 0 750 041 A2**

(12) **DEMANDE DE BREVET EUROPEEN**

(43) Date de publication:  
**27.12.1996 Bulletin 1996/52**

(51) Int Cl.<sup>6</sup>: **C12N 15/49, C07K 14/16,  
C12Q 1/68**

(21) Numéro de dépôt: **96108720.2**

(22) Date de dépôt: **15.01.1988**

(84) Etats contractants désignés:  
**AT BE CH DE ES FR GB GR IT LI LU NL SE**

(30) Priorité: **16.01.1987 US 3764**  
**11.02.1987 FR 8701739**  
**15.04.1987 FR 8705398**

(62) Numéro de dépôt de la demande initiale en  
application de l'article 76 CBE: **88400084.5**

(71) Demandeurs:  
• **INSTITUT PASTEUR**  
**F-75015 Paris (FR)**  
Etats contractants désignés:  
**BE CH DE ES GB GR IT LI LU NL SE AT**  
• **CENTRE NATIONAL DE**  
**LA RECHERCHE SCIENTIFIQUE**  
**75007 Paris (FR)**  
Etats contractants désignés:  
**FR**

(72) Inventeurs:  
• **Alizon, Marc**  
**75005 Paris (FR)**  
• **Montagnier, Luc**  
**92350 Le Plessis Robinson (FR)**

- **Guetard, Denise**  
**75015 Paris (FR)**
- **Clavel, Françoise**  
**Rockville, MD 20852 (US)**
- **Sonigo, Pierre**  
**75015 Paris (FR)**
- **Guyader, Mireille**  
**75017 Paris (FR)**
- **Tlollals, Pierre**  
**75013 Paris (FR)**
- **Chakrabarti, Lisa**  
**750 Paris (FR)**
- **Desrosiers, Ronald**  
**Hudson, Massachusetts 01749 (US)**

(74) Mandataire: **Desaix, Anne**  
**Ernest Gutmann - Yves Plasseraud S.A.**  
**3, rue Chauveau-Lagarde**  
**75008 Paris (FR)**

Remarques:

Cette demande a été déposée le 31 - 05 - 1996  
comme demande divisionnaire de la demande  
mentionnée sous le code INID 62.

(54) **Séquences de nucléotides de HIV-2**

(57) L'invention concerne une séquence de nucléotides caractérisée en ce qu'elle répond à la séquence nucléotidique représentée à la figure 1B ou à la figure 1C, ou en ce qu'elle contient la séquence nucléotidique représentée à la figure 1B ou à la figure 1C, ou en ce qu'il s'agit d'une partie de la séquence représentée à la

figure 1B ou à la figure 1C, ladite partie de séquence codant pour un peptide reconnu par des anticorps présents dans le sérum d'un patient infecté par un rétrovirus HIV-2 ou étant utilisable comme sonde pour la détection dans un échantillon biologique, de la présence d'un rétrovirus HIV-2.

**EP 0 750 041 A2**

## Description

La présente invention est relative à des peptides ayant des propriétés immunologiques, le cas échéant immunogènes, en commun avec des antigènes susceptibles d'être obtenus sous une forme purifiée, à partir de virus capables de provoquer des lymphadénopathies susceptibles d'être relayées ensuite par le syndrome d'immunodéficience acquise (SIDA) chez l'homme.

L'invention concerne en particulier des peptides antigéniques susceptibles d'être reconnus par des anticorps induits chez l'homme par des virus désignés par l'abréviation HIV, selon la nomenclature définie dans NATURE. Elle concerne également des peptides ayant des propriétés immunogènes ou susceptibles d'être rendus immunogènes *in vivo*, cette immunogénicité étant susceptible de se manifester par l'induction *in vivo* d'anticorps reconnaissant des antigènes caractéristiques des virus HIV-2 et même, au moins en ce qui concerne certains de ces peptides, des antigènes issus de HIV-1.

L'invention concerne en outre des applications de ces peptides à la fabrication de compositions pour le diagnostic *in vitro* chez l'homme de potentialité de certaines formes du SIDA et, en ce qui concerne certains d'entre eux, à la production de compositions immunogènes et de compositions vaccinales contre les rétrovirus HIV.

De même l'invention concerne les applications aux mêmes fins des anticorps susceptibles d'être induits *in vivo* par les peptides immunogènes ou rendus immunogènes et, pour certains de ces anticorps, leurs applications à la production de principes actifs de médicaments contre ces SIDA humains.

L'invention concerne également la mise en oeuvre de certains de ces peptides dans des procédés pour le diagnostic *in vitro* chez l'homme de certaines formes du SIDA, ainsi que leur application à la constitution de trousse ou "kits" de diagnostic.

Un premier rétrovirus dénommé LAV-1 ou HIV-1 a été isolé et décrit dans la demande de brevet GB.83/24.800 et une demande EP.84/401.834 du 14/09/84. Ce virus a également été décrit par F.Barre Sinoussi et al. dans Science, 220 n° 45-99, 20 pages 868-871.

Des variants de ce virus HIV-1 désignés par LAV ELI et LAV MAL, ont également été isolés, caractérisés et décrits dans la demande de brevet EP.84/-401.834.

Les virus HIV-1 et leurs variants possèdent les propriétés suivantes :

- ils ont pour cibles préférentielles les cellules Leu3 (ou lymphocytes T4) humaines et leurs cellules dérivées "immortalisées".
- ils ont une activité transcriptase inverse nécessitant la présence d'ions  $Mg^{2+}$  et présentent une forte activité pour le poly(adenylate-oligo-deoxythymidylase) poly(A)-oligo(dT)12-18)
- ils ont une densité de 1,16 à 1,17 sur gradient de sucrose,
- ils ont un diamètre moyen de 139 nanomètres et un noyau de diamètre moyen de 41 nanomètres,
- les lysats de ces virus contiennent une protéine p25 (protéine du noyau) qui ne croise pas immunologiquement avec la protéine p24 de HTLV-1,
- ils contiennent une protéine p42 appartenant à leur enveloppe,
- ils contiennent également une glycoprotéine d'enveloppe gp110 d'un poids moléculaire de 110.000.

L'isolement et la caractérisation de rétrovirus appartenant à une classe distincte et n'ayant qu'une parenté immunologique réduite avec les précédents, ont été décrits dans la demande de brevet européen n° 87/400.151.4. Ces rétrovirus qui ont été regroupés sous la désignation HIV-2, ont été isolés chez plusieurs malades africains présentant des symptômes d'une lymphadénopathie ou d'un SIDA.

Les rétrovirus du type HIV-2 comme les rétrovirus du type HIV-1, se caractérisent par un tropisme pour les lymphocytes T4 humains et par un effet cytopathogène à l'égard de ces lymphocytes, lorsqu'ils s'y multiplient, pour alors causer soit des poly-adénopathies généralisées et persistantes, soit un SIDA.

Plus généralement les rétrovirus purifiés par HIV-2 possèdent en général les propriétés suivantes :

- la cible préférentielle des rétrovirus HIV-2 est constituée par les cellules Leu3 (ou lymphocytes T4) humaines et pour des cellules "immortalisées" dérivées de ces lymphocytes T4 ;
- ils sont cytotoxiques pour les lymphocytes T4 humains
- ils ont une activité de transcriptase inverse nécessitant la présence d'ions  $Mg^{2+}$  et présentant une forte activité pour le poly(adenylate-oligodéoxythymidylase) (poly(A)-oligo(dT) 12-18) ;
- ils ont une densité de 1,16 dans un gradient de sucrose ;
- ils ont un diamètre moyen de 140 nanomètres et un noyau ayant un diamètre moyen de 41 nanomètres ;
- ils peuvent être cultivés dans des lignées permanentes du type HUT ou exprimant la protéine T4 ;
- ils ne sont pas infectieux pour les lymphocytes T8 ;
- les lysats de ces virus contiennent une protéine p26 qui ne croise pas immunologiquement avec la protéine p24

du virus HTLV-I ou du virus HTLV-II ;

- ces lysats contiennent en outre une protéine p16 qui n'est pas reconnue immunologiquement par la protéine p19 de HTLV-I ou de HTLV-II dans des essais de radioimmuno-précipitation ;
- ils contiennent en outre une glycoprotéine d'enveloppe ayant un poids moléculaire de l'ordre de 130.000-140.000 qui ne croise pas immunologiquement avec la gp110 des HIV-1, mais qui en revanche croise immunologiquement avec la glycoprotéine d'enveloppe gp140 de STLV-III (virus isolé chez le singe) ;
- ces lysats contiennent encore des antigènes marquables par la <sup>35</sup>S-cystéine, dont les poids moléculaires s'étagent entre 32.000 et 42.000-45.000 : ils comprennent notamment un antigène ayant un poids moléculaire de l'ordre de 36.000 et un antigène ayant un poids moléculaire de l'ordre de 42.000, l'un de ces antigènes (p36 et p42) constituant vraisemblablement une glycoprotéine transmembranaire du virus HIV-2 ;
- l'ARN génomique des HIV-2 n'hybride pas avec l'ARN génomique de HIV-1 dans des conditions stringentes ;
- dans des conditions non stringentes, l'ARN génomique de HIV-2 n'hybride, ni avec le gène env et le LTR qui le jouxte, de HIV-1, ni avec des séquences de la région pol du génome de HIV-1 ;
- dans des conditions non stringentes, il hybride faiblement avec des séquences de nucléotides de la région de HIV-1.

Un autre rétrovirus dénommé SIV-1, cette dénomination remplaçant la dénomination antérieurement connue STLV III, a été isolé chez le singe macaque rhésus. (M.D.Daniel et al. Science 228, 1201 (1985) N.L.Letwin et al, Science 230, 71 (1985) sous l'appellation "STLV-III<sub>mac</sub>").

Un autre rétrovirus, désigné "STLV-III<sub>AGM</sub>", (ou SIV<sub>AGM</sub>) a été isolé chez des singes verts sauvages. Mais, contrairement au virus présent chez le singe macaque rhésus, la présence de "STLV-iii<sub>AGM</sub>" ne semble pas induire une maladie du type SIDA chez le singe vert d'Afrique.

Une souche du rétrovirus SIV-1<sub>mac</sub> a été déposée à la CNCM le 7 Février 1986 sous le n° I-521. Des études ont montré que le rétrovirus SIV-1 comporte certaines protéines possédant une certaine parenté immunologique avec des protéines ou glycoprotéines structurales susceptibles d'être obtenues dans des conditions analogues, à partir de HIV-2. Ce rétrovirus SIV-1, dont on a constaté le caractère infectieux chez les singes, avait été désigné par STLVIII par les chercheurs qui l'ont isolé (références bibliographiques précitées).

Pour la commodité du langage, ces virus ne seront plus désignés dans ce qui suit que par l'expression SIV (l'expression SIV est l'abréviation anglaise de "Simian Immunodeficiency Virus" (virus d'immunodéficience du singe)) éventuellement suivi d'une abréviation désignant l'espèce de singe dont ils sont issus, par exemple, MAC (ou mac) pour le macaque ou AGM pour le singe vert d'Afrique (abréviation de "African Green Monkey").

En mettant en oeuvre les mêmes techniques que celles rappelées plus haut, il a été constaté que l'on pouvait également obtenir à partir de SIV-1<sub>mac</sub> :

- une protéine principale du noyau p27, ayant un poids moléculaire de l'ordre de 27 kilodaltons,
- une glycoprotéine majeure d'enveloppe, gp140,
- une protéine vraisemblablement transmembranaire p32, qui n'est guère observée en RIPA lorsque le virus a au préalable été marqué par la <sup>35</sup>S-cystéine, mais qui peut être observée dans les essais d'immunoempreintes (Western blots), sous forme de bandes larges.

Des études plus précises ont été réalisées en ce qui concerne les précédents virus HIV-2 et SIV. La poursuite de l'étude des rétrovirus HIV-2 a également conduit à l'obtention de séquences d'ADN complémentaires (ADNc) des ARNs de leurs génomes. La séquence nucléotidique complète de l'ADNc d'un rétrovirus représentatif de la classe HIV-2 (HIV-2 ROD) a été déposée le 21/02-1986 à la CNCM sous le n° I-522, sous le nom de référence LAV-II ROD).

Cette séquence nucléotidique et les phases de lecture ouverte qu'elle contient sont indiqués à la figure 1 A.

En outre, la poursuite de l'étude d'autres rétrovirus a également permis d'aboutir à l'obtention de leurs séquences nucléotidiques complètes. Il en est en particulier ainsi de l'ADNc dérivé de l'ARN génomique de SIV.

Le clonage et le séquençage du virus SIV-1<sub>mac</sub> qui ont permis l'obtention de sa séquence nucléotidique ont été réalisés dans les conditions suivantes :

L'ADN de cellules HUT 78 infectées par le virus SIV (isolat STLV-III<sub>mac</sub> 142-83 décrit par Daniel et al.(1985) Science, 228, p.1201-1204, digéré partiellement par l'enzyme de restriction Sau3A a été cloné au site BamHI du bactériophage vecteur Lambda ELBL3 pour constituer une banque génomique. Les 2 millions de phages recombinants de la banque génomique ainsi constituée ont été criblés in situ en conditions de sécurité P3, à l'aide de séquences du virus HIV2 provenant des clones lambda-ROD4, lambda-ROD35 et E2 (Clavel et al. (1986-Nature, 324, p.691.) et nick-translatées.

L'hybridation a été réalisée en 5xSSC à 50°C et les lavages en 2xSSC à 5°C. Un seul clone contenant l'ensemble des séquences virales a été obtenu. Ce clone est désigné par lambda-SIV-1. L'insérat du phage lambda-SIV-1 mesure 16,5 kb au total et comprend un provirus intégré auquel manquent seulement les 250 premières bases du LTR gauche,

alors que le LTR droit est complet.

Le provirus intégré a été séquencé par la méthode des didéoxynucléotides après sous-clonage de fragments aléatoires dans le phage M13mp8. 300 sous-clones ont été analysés.

Des fragments d'ADNc provenant du clone Lambda SIV-1 insérés dans des plasmides pSIV-1.1 et pSIV-1.2 ont été déposés à la CNCM le 15 Avril 1987, sous les numéros I-658 (pSIV-1.1) et I-659 (pSIV-1.2).

Les résultats ont été mentionnés dans les figures décrites ci-après.

La figure 1B représente la séquence nucléotidique du génome viral de SIV et les séquences qui en sont déduites pour les protéines virales correspondant aux produits des gènes gag, pol, env, Q, X, R, tat, art, F.

Les figures 3 à 11 et la figure 1C représentent les comparaisons des produits théoriques des gènes viraux et des LTR entre HIV2 et SIVmac. (λSIV-1).

L'invention concerne de plus les fragments d'ADNc déduits de l'ADNc issu du génome entier de SIV-1, ces fragments contenant une ou plusieurs séquences issues de la séquence complète d'ADNc et qui codent pour des peptides intéressants de l'invention. Ces séquences sont indiquées à la figure 1B et, à la figure 1C pour ce qui a trait à la séquence LTR du virus,

Les séquences nucléiques de l'ADNc de SIV ont été placées en correspondance avec les séquences nucléiques du virus HIV-2 ROD pour ce qui concerne la séquence LTR (figure 1C). Cette présentation que l'on retrouve pour le génome entier en rapprochant la figure 1B des figures 3 à 11 permet de repérer ou de déduire les acides nucléiques ayant des éléments de structure essentiels communs aux deux virus.

L'invention concerne naturellement aussi l'utilisation des cADNs issus de SIV ou de leurs fragments (ou de recombinaisons les contenant) en tant que sondes, pour le diagnostic de la présence ou non de virus HIV-2 dans des échantillons de sérums ou d'autres liquides ou tissus biologiques obtenus à partir de patients suspectés d'être porteurs du virus HIV-2. Ces sondes sont de préférence marquées également (marqueurs radio-actifs, enzymatiques, fluorescents, etc.). Des sondes particulièrement intéressantes pour la mise en oeuvre du procédé de diagnostic du virus HIV-2 ou d'un variant de HIV-2 peuvent être caractérisées en ce qu'elles comprennent la totalité ou une fraction de l'ADNc complémentaire du génome du virus SIV ou encore notamment les fragments recombinants contenus dans divers clones.

Les sondes mises en oeuvre dans ce procédé de diagnostic du virus HIV-2 et dans les kits de diagnostic ne sont en aucune façon réduites aux sondes décrites précédemment. Elles comprennent au contraire toutes les séquences nucléotidiques issues du génome du virus SIV, d'un variant de SIV ou d'un virus proche par sa structure, dès lors qu'elles permettent la détection dans des fluides biologiques de personnes susceptibles de développer un SIDA, d'anticorps dirigés contre un HIV-2 ou d'un virus qui en est proche.

La détection peut être réalisée de toutes façons en soi connues. Elle peut comprendre une mise en contact de ces sondes soit avec les acides nucléiques obtenus à partir des cellules contenues dans ces sérums ou autres milieux biologiques, par exemple liquides céphalo-rachidiens, salives, etc... Elle peut aussi comprendre une mise en contact de ces sondes avec ces milieux eux-mêmes dès lors que leurs acides nucléiques ont été rendus accessibles à l'hybridation avec ces sondes, et ce dans des conditions permettant l'hybridation entre ces sondes et ces acides nucléiques. L'étape finale du diagnostic *in vitro* comprend alors la détection de l'hybridation éventuellement produite. Le susdit diagnostic mettant en jeu des réactions d'hybridation peut également être réalisé à l'aide de mélanges de sondes respectivement originaires d'un HIV-2 et d'un SIV-1 ou d'un HIV-1, d'un HIV-2 et d'un SIV, dès lors qu'il n'est pas nécessaire de faire une différence entre le type de virus recherché.

D'une façon générale, le procédé de diagnostic de la présence ou non du virus HIV-2 ou d'un variant dans des échantillons de sérums ou d'autres liquides ou tissus obtenus à partir de patients suspectés d'être porteurs du virus HIV-2 comprend les étapes suivantes :

- 1/ au moins une étape d'hybridation conduite dans des conditions stringentes, par mise en contact de l'ADN de cellules de l'échantillon du patient suspect avec l'une des susdites sondes marquées sur une membrane appropriée,
- 2/ le lavage de ladite membrane avec une solution assurant la conservation de ces conditions stringentes de l'hybridation,
- 3/ la détection de la présence ou non du virus HIV-2 par une méthode d'immunodétection.

Dans un autre mode de réalisation préféré du procédé selon l'invention l'hybridation précitée est conduite dans des conditions non stringentes et le lavage de la membrane est réalisé dans des conditions adaptées à celles de l'hybridation.

Il va de soi que l'invention concerne les acides nucléiques correspondant à des séquences placées en des régions analogues de variants de SIV ainsi que tous les acides nucléiques dont les modifications résulteraient de la mise à profit de la dégénérescence du code génétique.

Les études comparatives qui ont aussi permis d'aboutir à des résultats relatifs aux protéines de noyau (core), ci-

après dénommées "protéines gag" et aux protéines d'enveloppes, ci-après dénommées "protéines env", ont également été rapportés dans la demande de brevet européen n° 87/400.151.4, déjà citée. Ces résultats montrent que les protéines du noyau (protéines gag) dans HIV-2 présentent des différences moins accentuées par rapport à celles des virus HIV-1, que les protéines d'enveloppe (protéines env). Globalement les protéines env dans HIV-2 se sont révélées présenter des parentés immunologiques extrêmement faibles, sinon inexistantes, avec les protéines env correspondantes des virus HIV-1.

Au contraire des études comparatives effectuées entre les structures des séquences d'ADNc des virus HIV-2 et SIV permettent de mettre en évidence certaines caractéristiques communes qui apparaissent au niveau des protéines. Globalement, les protéines de HIV-2 et de SIV-1 montrent des parentés immunologiques importantes.

La glycoprotéine majeure d'enveloppe de HIV-2 s'est révélée être plus proche immunologiquement de la glycoprotéine majeure d'enveloppe de SIV que de la glycoprotéine majeure d'enveloppe de HIV-1.

Ces constatations s'imposent non seulement au niveau des poids moléculaires : 130-140 kilodaltons pour les glycoprotéines majeures de HIV-2 et de SIV contre environ 110 pour la glycoprotéine majeure d'enveloppe de HIV-1, mais aussi au niveau des propriétés immunologiques, puisque des sérums prélevés à partir de malades infectés par HIV-2, et plus particulièrement des anticorps formés contre la gp140 de HIV-2 reconnaissent la gp140 de SIV-1mac, alors que dans des essais semblables les mêmes sérums et les mêmes anticorps de HIV-2 ne reconnaissent pas la gp110 de HIV-1. Mais les sérums anti-HIV-1 qui n'ont jamais réagi avec la gp140 de HIV-2 précipitent une protéine de 26 Kdal marquée par la <sup>35</sup>S-cystéine, contenue dans les extraits de HIV-2.

La protéine majeure du noyau (core) de HIV-2 semble présenter un poids moléculaire moyen (environ 26.000) intermédiaire entre celui de la p25 de HIV-1 et la p27 de SIV.

Ces observations résultent des essais réalisés avec des extraits viraux obtenus à partir du HIV-2 isolé à partir de l'un des patients susmentionnés. Des résultats similaires ont été obtenus avec des extraits viraux du HIV-2 isolé à partir du second patient.

Des études plus poussées ont conduit les inventeurs à reconnaître une première classe de peptides ayant des séquences d'acides aminés soit identiques, soit proches de séquences contenues à l'intérieur des structures des protéines gag et env de HIV-2 ou de SIV voire de HIV-1. Ces peptides sont notamment applicables au diagnostic d'une infection chez l'homme par le virus HIV-2 ou de l'un de ses variants.

A cet égard la présente invention concerne également des procédés et des compositions de diagnostic pour la détection *in vitro* d'anticorps dirigés contre un virus HIV-2 ou de ses variants, plus particulièrement dans des échantillons biologiques, notamment des sérums de patients ayant subi une infection par le virus HIV-2, certains de ces peptides permettant une discrimination particulièrement poussée entre les infections dues à des virus HIV-2 et à des virus HIV-1.

Ces études poussées ont également conduit à la possibilité de synthétiser des peptides immunogènes ou susceptibles d'être rendus immunogènes, présentant des caractéristiques de structures leur permettant d'induire *in vivo* la production d'anticorps susceptibles de reconnaître des protéines env à la fois dans HIV-1 et dans HIV-2 et, au moins pour certains de ces peptides, de se fixer tant sur des virus HIV-1 que sur des virus HIV-2, plus particulièrement aux fins de les neutraliser. L'utilisation de ces derniers types de peptides est donc particulièrement indiquée pour la production de principes actifs de vaccins contre les virus HIV, donc contre le SIDA.

Pour désigner ci-après les résidus d'acides aminés entrant dans la constitution des peptides selon l'invention, on aura recours, pour ceux des acides aminés ayant une signification univoque à la nomenclature internationale désignant chaque acide aminé naturel par une lettre unique (lettre majuscule) selon le tableau des correspondances qui suit :

M Méthionine  
L Leucine  
I Isoleucine  
V Valine  
F Phénylalanine  
S Sérine  
P Proline  
T Thréonine  
A Alanine  
Y Tyrosine  
H Histidine  
Q Glutamine  
N Asparagine  
K Lysine  
D Acide Aspartique  
E Acide glutaminique  
C Cystéine

## EP 0 750 041 A2

W Tryptophane  
R Arginine  
G Glycine

5        Lorsqu'un acide aminé pourra, en raison de sa position au sein de la chaîne d'acides aminés caractéristique d'un peptide déterminé, prendre plusieurs significations, il pourra soit être désigné par un tiret "-", si sa signification peut être quelconque, soit par une lettre minuscule lorsque cet acide aminé pourra présenter un nombre limité de significations préférées, ce nombre étant cependant toujours supérieur à 1. Dans ce dernier cas, les significations possibles de cette lettre minuscule seront toujours précisées en rapport avec le peptide auquel il appartient.

10       Afin de faciliter la lecture, ces peptides seront désignés par une abréviation env ou gag suivie d'un indice numérique, par référence à des séquences d'acides aminés contenues, selon le cas, soit dans les protéines env soit dans les protéines gag de certains HIV-1, HIV-2 ou SIV. Il y sera encore fait référence dans ce qui suit.

Enfin dans les définitions qui suivent

15       - les groupes X représentent soit un groupe NH<sub>2</sub> libre ou amidé, notamment par un ou deux groupes alcoyle comprenant de 1 à 5 atomes de carbone, soit un groupe peptidique comprenant de 1 à 5 acides aminés, dont l'acide aminé N-terminal présente lui-même un groupe NH<sub>2</sub> libre ou amidé comme précédemment indiqué, et  
- les groupes Z représentent, soit un groupe -OH libre ou alcoyle et contenant alors un groupe alcoyle comprenant de 1 à 5 atomes de carbone, soit un groupe peptidique comprenant de 1 à 5 acides aminés, dont l'acide aminé C-terminal présente lui-même un groupe -OH libre ou alcoyle, comme précédemment indiqué, les groupes de 1 à 5 acides aminés le cas échéant contenus dans X ou Z ou dans les deux à la fois étant tels, que leur présence n'est pas incompatible avec la préservation pour l'essentiel des propriétés immunologiques, le cas échéant immunogènes, des peptides qui en sont dépourvus.

25       Les peptides selon l'invention, qui ont en commun des propriétés immunologiques avec des antigènes de HIV-2 et, pour certains d'entre eux également avec des antigènes de HIV-1 ou de ses variants, sont caractérisés en ce qu'ils ont également une structure peptidique en commun avec les antigènes de SIV. De façon avantageuse, ces peptides comprennent normalement au plus 40 résidus d'acides aminés.

Des peptides préférés sont les suivants :

30

35

40

45

50

55



env1

XRV-AIEKYL-DQA-LN-WGCAFRQVCZ

5

env2

X-LE-AQI-QQEKNNMYELQKLNZ

env3

XELGDYKLVEITPIG-APT--KR-----Z

10

env4

X-----VTV-YGVP-WK-AT--LFCA-Z

env5

15

X---QE--L-NVTE-F--W-NZ

env6

XL---S-KPCVKLTPLCV--Z

20

env7

X---N-S-IT--C-K-----Z

env8

X-I---YC-P-G-A-L-C-N-TZ

25

env9

X-----A-C-----W--Z

env10

30

X-G-DPE-----NC-GEF-YCN-----NZ

35

env11

X-----C-IKQ-I-----G---YZ

40

Plus particulièrement l'invention concerne les peptides suivants :

45

50

55

env1  
XRV-AIEKYL-DQA-LN-WGCAFRQVCZ

5 env2  
X-LE-AQIQQEKKNMYELQKLNZ

env3  
10 XELGDYKLVEITPIG-APT--KR-----Z

env4  
X----VTV-YGVP-W--AT--LFCA-Z

env5  
15 X----E--L-NVTE-F--W-NZ

env6  
XL---S-KPCVKL-PLC---Z

20 env7  
X---N-S-I---C-K----Z

env8  
25 X-I---YC-P-G-A-L-C-N-TZ

env9  
X-----A-C-----W--Z

30 env10  
X-G-DPE-----NC-GEF-YC-----NZ

env11  
35 X-----C-I-Q-I-----G---YZ

Des peptides avantageux correspondant aux précédents, présentent les formules qui suivent :

env1  
40 XRVTAIEKYLQDQARLNSWGCAFRQVCZ, ou  
XRVTAIEKYLKDQAQLNAWGCAFRQVCZ

env2  
45 XSLEQAQIQQEKKNMYELQKLNSWZ, ou  
XLLEEAQIQQEKKNMYELQKLNSWZ

env3  
50 XELGDYKLVEITPIGFAPTKEKRYSSAHZ, ou  
XELGDYKLVEITPIGLAPTNVKRYTTG-Z

55 (On remarquera que les peptides env1, env2, env3 attestent de la très grande parenté entre HIV-2 et SIV-1. En effet le premier peptide est inclu dans le génome de HIV-2 et le second, dans celui de SIV-1).

**env4**

**XabcdVTVeYGVpfWogATHiLFCAjZ,**

5 dans lesquels les lettres de a à j peuvent avoir les significations suivantes :

a est C, E ou D  
b est T, K, D, N ou I  
c est Q ou L  
10 d est Y ou W  
e est F ou Y  
f est T, V ou A  
g est N ou E  
h est I ou T  
15 i est P ou T  
j est T ou S  
o est K ou R

**env5**

**XabcoEdeLfNVTEgFhiWjNZ,**

20 dans lequel les lettres de a à j peuvent avoir les significations suivantes :

a est D ou P  
b est D ou N  
c est Y ou P  
d est I, V, I ou L  
e est T, V, E ou A  
30 f est V, G ou E ou -  
g est A, N, G ou S  
h est D ou N  
i est A ou M  
j est N, K ou E  
35 o est Q ou S

**env6**

**XLabcSdKPCVKLoPLCuefKZ,**

40 dans lequel les lettres de a à f peuvent avoir les significations suivantes :

a est F ou W  
45 b est E ou D  
c est T ou Q  
d est I ou L  
e est A, S ou T  
f est M ou L  
50 o est T ou S  
u est V ou I

**env7**

**XabCNxSyIocdCeKfghiZ,**

55 dans lequel les lettres de a à i et x et y peuvent avoir les significations suivantes :  
a est N ou T ou I

EP 0 750 041 A2

b est H ou S ou N  
c est E ou Q  
d est S, A ou C  
e est D ou P  
5 f est H, V ou D  
g est Y ou S  
h est W ou F  
i est D ou E  
x est T ou R  
10 y est V ou A  
o est T ou Q

env8

15 **XaIbcdYCxPeGfAgLhCiNjTz,**

dans lequel les lettres de a à k et x peuvent avoir les significations suivantes :

a est A ou P  
20 h est R ou P  
c est F, I ou C  
d est R ou H  
e est P ou A  
f est Y ou F  
25 g est L ou I  
h est R ou K  
i est - ou N  
j est D ou K  
x est A ou T  
30

env9

**XwabcxyAdCefghizWjkZ,**

35 dans lequel les lettres de a à k et x à z peuvent avoir les significations suivantes :

a est K ou - ou E  
b est R ou -  
c est P ou M ou I  
40 d est W ou H ou Y  
e est W ou N ou T ou R  
f est F ou I  
g est K ou S ou N ou G  
h est G ou R ou E  
45 i est - ou A ou T  
j est K ou N ou D ou S  
k est D ou A ou N ou K ou E  
w est N, D ou I  
x est R ou G ou K  
50 y est Q ou K ou R  
z est K ou E ou Q ou N

env10

55 **XaGbDPEcdefghNCiGEFjYCokxlmnNZ,**

dans lequel les lettres de a à n et x peuvent avoir les significations suivantes :

## EP 0 750 041 A2

a est K ou - ou G  
 b est S ou G ou -  
 c est V ou I  
 d est A ou V ou T  
 5 e est Y ou T ou M ou F  
 f est M ou H  
 g est W ou S  
 h est T ou F  
 i est R ou G  
 10 j est L ou F  
 o est N ou K  
 k est M ou S  
 l est W ou Q ou K ou G  
 m est F ou L  
 15 n est L ou F  
 x est T ou S ou N

### env11

20 **XabcdwCeIoQfIxgyhizGjklYZ,**

dans lequel les lettres de a à 1 et w à z peuvent avoir les significations suivantes :

a est R ou T ou S ou N  
 25 b est N ou I  
 c est Y ou T  
 d est A ou L ou V  
 e est H ou R  
 f est I ou F  
 30 g est T ou M  
 h est H ou Q ou A  
 i est K ou E  
 j est R ou K  
 k est N ou A  
 35 l est V ou M  
 w est P ou Q  
 x est N ou K  
 y est W ou V  
 z est V ou T ou K  
 40 o est K ou R

La structure du peptide antigénique codé par le gène gag et désigné par gag1 est également représentée ci-après :

45 **XDCKLVLKGLGaNPTEEMLTaz,**

dans lequel la lettre a désigne M ou T.

Il sera remarqué que, d'une façon générale, les aminoacides ayant une signification univoque (donc représentés par une lettre majuscule correspondant à la nomenclature internationale) qui interviennent dans les définitions qui précèdent des peptides selon l'invention, se trouvent être la correspondance avec des aminoacides identiques placés dans le même ordre dans les séquences env ou gag correspondantes de la protéine env ou gag d'au moins l'un des HIV, ou de SIV-1.

Les positions de ces séquences sont soulignées et repérées au sein des séquences d'acides aminés des protéines env respectivement de HIV-2 ROD (CNCM n° I-532) et HIV-1 BRU (CNCM n° I-232) représentées à la figure 2. Par ailleurs, les alignements des acides aminés des protéines env et gag respectivement de SIV-1mac (CNCM n° I.521) et de HIV-2 ROD sont présentées à la figure 3 et à la figure 4.

Les traits pleins qui apparaissent en certaines localisations de ces séquences visent à souligner que certains aminoacides contenus dans ces séquences ont été volontairement délégués au plan de la présentation, afin de permettre

## EP 0 750 041 A2

la mise en alignement d'acides aminés respectivement identiques (alors marqués d'un astérisque) ou de deux points verticaux sur une même ligne verticale dans les séquences des protéines correspondantes de HIV-1 et de HIV-2 d'une part, de SIV et de HIV-2 d'autre part.

5 Outre les peptides précités, l'invention concerne également les peptides modifiés par insertion et/ou délétion et/ou substitution d'un ou plusieurs acides aminés, pour autant que les propriétés antigéniques ou immunogènes desdits peptides ne sont pas modifiées, ou que les propriétés de reconnaissance de l'antigène ou de l'anticorps avec lesdits peptides ne sont pas substantiellement modifiées.

Dans un mode de réalisation particulièrement préféré, l'invention concerne des peptides ayant des propriétés immunologiques en commun avec l'ossature peptidique de la glycoprotéine d'enveloppe des virus de la classe HIV-2, ces peptides contenant un nombre de résidus d'acides aminés n'excédant pas 40.

Ces peptides préférés selon l'invention ont les séquences suivantes :

### env1

15 RVT A I E K Y L Q D Q A R L N S W G C A F R Q V C  
A I E K Y L Q D Q  
R V S A I E K Y L K D Q A Q L N A W G C A F R Q V C  
20 A I E K Y L K D Q

### env2

S L E Q A Q I Q Q E K N M Y E L Q K L N S W  
25 Q I Q Q E K N  
L L E E A Q I Q Q E K N M Y E L Q K L N S W

### env3

30 E L G D Y K L V E I T P I G F A P T K E K R Y S S A H  
Y K L V E I T P I G F A P T K E K  
E L G D Y K L V E I T P I G L A P T N V K R Y T T G -  
35 Y K L V E I T P I G L A P T N V K

### env4

40 C T Q Y V T V F Y G V P T W K N A T I P L F C A T  
V T V F Y G V P T W K N A T  
C I Q Y V T V F Y G V P A W R N A T I P L F C A T  
45 V T V F Y G V P A W R N A T

E K L W V T V Y Y G V P V W K E A T T T L F C A S  
50 V T V Y Y G V P V W K E A T  
E D L W V T V Y Y G V P V W K E A T T T L F C A S  
V T V Y Y G V P V W K E A T  
55 D N L W V T V Y Y G V P V W K E A T T T L F C A S  
V T V Y Y G V P V W K E A T

EP 0 750 041 A2

env5

DDYQEITL-NVTEAFDAWNN  
5 L-NVTEAF  
DDYSELAL-NVTESEDAWEN  
L-NVTESEF  
10 PNPQEVVLNVNTENFNMWKN  
LVNVNTENF  
PNPQEIELENVTEGFNMWKN  
LENVTEGF  
15 PNPQEIALENVNTENFNMWKN  
LENVNTENF

env6

ETSIKPCVKLTPLCVAMK  
ETSIKPCVKLSPLCITMR  
25 DQSLKPCVKLTPLCVSLK  
DQSLKPCVKLTPLCVTLN  
PCVKLTPLCV

env7

NHCNTSVITESCD  
NTSVIT  
35 NHCNTSVIQECCD  
NTSVIQ  
TSCNTSVITQACP  
40 NTSVIT  
INCNTSVITQACP  
NTSVIT  
45 INCNTSAITQACP  
NTSAIT

env8

50 YCAPPGYALLRC-NDT  
YCAPAGFAILKCNNKT  
YCAPAGFAILKCNDKK  
55 YCAPAGFAILKCRDKK

env9

5 NKRPRQAWCWFKG-KWKD  
 NERPKQAWCRFGG-NWKE  
 N--MRQAHCNISRAKWNA  
 D--IRRAYCTINETEWDK  
 10 I--IGQAHCNISRAQWSK

env10

15 KGSDPEVAYMWTNCRGEFLYCNMTWFLN  
 NCRGEFLYCN  
 GG-DPEVTFMWTNCRGEFLYCKMNWFLN  
 NCRGEFLYCK  
 20 -GGDPEIVTHSFNCGGEFFYCNSTQLFN  
 NCGGEFFYCN  
 -GGDPEITTHSFNCRGEFFYCNSTSKLFN  
 NCRGEFFYCN  
 25 -GGDPEITTHSFNCGGEFFYCNSTGLFN  
 NCGGEFFYCN

env11

30 RNYAPCHIKQIINTWHKVGGRNVY  
 CHIKQII  
 35 RNYVPCHIRQIINTWHKVGKENVY  
 CHIRQII  
 TITLPCRIKQFINMWQEVGKAMY  
 40 CRIKQFI  
 SITLPCRIKQIINMWQKTCKAMY  
 CRIKQII  
 45 NITLQCRIKQIIKMOVAGR-KAIY  
 CRIKQII

gag1

50 DCKLVLKGLGTNPTLEEMLT A

Les peptides selon l'invention peuvent encore avantageusement être préparés par les techniques classiques, dans le domaine de la synthèse des peptides. Cette synthèse peut être réalisée en solution homogène ou en phase solide.

55 Par exemple, on aura recours à la technique de synthèse en solution homogène décrit par HOUBENWEYL dans l'ouvrage intitulé "Méthode der Organischen Chemie" (Méthode de la Chimie Organique) édité par E. Wunsch, vol. 15-I et II., THIEME, Stuttgart 1974.

Cette méthode de synthèse consiste à condenser successivement deux-à-deux les aminoacyles successifs dans



l'ordre requis, ou à condenser des aminoacyles et des fragments préalablement formés et contenant déjà plusieurs aminoacyles dans l'ordre approprié, ou encore plusieurs fragments préalablement ainsi préparés, étant entendu que l'on aura eu soin de protéger au préalable toutes les fonctions réactives portées par ces aminoacyles ou fragments, à l'exception des fonctions amines de l'un et carboxyles de l'autre ou vice-versa, qui doivent normalement intervenir dans la formation des liaisons peptidiques, notamment après activation de la fonction carboxyle, selon les méthodes bien connues dans la synthèse des peptides. En variante, on pourra avoir recours à des réactions de couplage mettant en jeu des réactifs de couplage classique, du type carbodiimide, tels que par exemple la 1-éthyl-3-(3-diméthyl-amino-propyl)-carbodiimide. Lorsque l'aminoacycle mis en œuvre possède une fonction acide supplémentaire (notamment dans le cas de l'acide glutamique), ces fonctions seront par exemple protégées, par des groupes t-butyloxyester.

Dans le cas de la synthèse progressive, l'acide aminé par acide aminé, la synthèse débute de préférence par la condensation de l'acide aminé C-terminal avec l'acide aminé qui correspond à l'acide aminé voisin dans la séquence désirée et ainsi de suite, de proche en proche, jusqu'à l'acide aminé N-terminal. Selon une autre technique préférée de l'invention, on a recours à celle décrite par R.D. MERRIFIELD dans l'article intitulé "Solid phase peptide synthesis" (J. Am. Soc., 45, 2149-2154).

Pour fabriquer une chaîne peptidique selon le procédé de MERRIFIELD, on a recours à une résine polymère très poreuse, sur laquelle on fixe le premier acide aminé C-terminal de la chaîne. Cet acide aminé est fixé sur la résine par l'intermédiaire de son groupe carboxylique et sa fonction amine est protégée, par exemple par le groupe t-butyloxy-carbonyl.

Lorsque le premier acide aminé C-terminal est ainsi fixé sur la résine, on enlève le groupe protecteur de la fonction amine en lavant la résine avec un acide.

Dans le cas où le groupe protecteur de la fonction amine est le groupe t-butyloxy-carbonyl, il peut être éliminé par traitement de la résine à l'aide d'acide trifluoroacétique.

On couple ensuite le deuxième acide aminé qui fournit le second amino-acycle de la séquence recherchée, à partir du résidu amino-acycle C-terminal sur la fonction amine déprotégée du premier acide aminé C-terminal fixé sur la chaîne. De préférence, la fonction carboxyle de ce deuxième acide aminé est activée, par exemple par la dicyclohexyl-carbodiimide, et la fonction amine est protégée, par exemple par le t-butyloxy-carbonyl.

On obtient ainsi la première partie de la chaîne peptidique recherchée, qui comporte deux acides aminés, et dont la fonction amine terminale est protégée. Comme précédemment, on déprotège la fonction amine et on peut ensuite procéder à la fixation du troisième amino-acycle, dans les conditions analogues à celles de l'addition du deuxième acide aminé C-terminal.

On fixe ainsi, les uns après les autres, les acides aminés qui vont constituer la chaîne peptidique sur le groupe amine chaque fois déprotégé au préalable de la portion de la chaîne peptidique déjà formée, et qui est rattachée à la résine.

Lorsque la totalité de la chaîne peptidique désirée est formée, on élimine les groupes protecteurs des différents acides aminés constituant la chaîne peptidique et on détache le peptide de la résine par exemple à l'aide d'acide fluorhydrique.

L'invention concerne également les oligomères hydrosolubles des peptides monomères sus-indiqués. L'oligomérisation peut provoquer un accroissement de l'immunogénicité des peptides monomères selon l'invention. Sans qu'une telle indication chiffrée puisse être considérée comme limitative, on mentionnera néanmoins que ces oligomères peuvent, par exemple, contenir de 2 à 10 unités monomères.

Les unités monomères entrant dans cet oligomère sont soit toutes constituées par le polypeptide de séquence 1 ou par le polypeptide de séquence 2, soit par l'un et l'autre de ces polypeptides.

On peut avoir recours, pour réaliser l'oligomérisation, à toute technique de polymérisation couramment utilisée dans le domaine des peptides, cette polymérisation étant conduite jusqu'à l'obtention d'un oligomère ou polymère contenant le nombre de motifs monomères requis pour l'acquisition de l'immunogénicité désirée.

Une méthode d'oligomérisation ou de polymérisation du monomère consiste dans la réaction de celui-ci avec un agent de réticulation tel que le glutaraldéhyde.

On peut également avoir recours à d'autres méthodes d'oligomérisation ou de couplage, par exemple à celle mettant en jeu des couplages successifs d'unités monomères, par l'intermédiaire de leurs fonctions terminales carboxyle et amine en présence d'agents de couplage homo- ou hétéro- bifonctionnels.

On peut également pour la production de molécules comportant un ou plusieurs motifs de 17 acides aminés tels que définis ci-dessus, avoir recours à des techniques du génie génétique mettant en œuvre des micro-organismes transformés par un acide nucléique déterminé comprenant des séquences nucléotidiques appropriées correspondantes.

L'invention concerne également les acides nucléiques contenant une ou plusieurs séquences issues de la séquence de l'ADNc du virus HIV-2 ROD. Ces séquences repérées par la numérotation figurant sur la séquence précédemment décrite, codent pour certains peptides intéressants de l'invention.

Séquence codant pour <u>env1</u> nucléotides 7850 à 7927				
	"	"	<u>env2</u>	" 8030 à 8095
5	"	"	<u>env3</u>	" 7601 à 7636
	"	"	<u>env4</u>	" 6170 à 6247
	"	"	<u>env5</u>	" 6294 à 6349
10	"	"	<u>env6</u>	" 6392 à 6445
	"	"	<u>env7</u>	" 6724 à 6763
	"	"	<u>env8</u>	" 6794 à 6838
	"	"	<u>env9</u>	" 7112 à 7162
15	"	"	<u>env10</u>	" 7253 à 7336
	"	"	<u>env11</u>	" 7358 à 7426
	"	"	<u>gag1</u>	" 1535 à 1597

20

L'invention concerne enfin les acides nucléiques correspondants du virus SiV, contenant une ou plusieurs séquences issues de l'ADNc du virus SIV-1. Ces séquences codant pour les peptides env1 à env11 et gag1 peuvent être repérés sur la figure 3 par comparaison avec les séquences correspondantes décrites pour HIV-2.

Il va de soi que l'invention concerne les acides nucléiques correspondant à des séquences placées en des régions analogues des ADNc dérivés de variants de HIV-2 ROD ou de SIV, ainsi que tous les acides nucléiques dont les modifications vis à vis des précédents résulteraient de la mise à profit de la dégénérescence du code génétique.

L'invention concerne encore les conjugués obtenus par couplage covalent des peptides selon l'invention (ou des susdits oligomères) à des molécules porteuses (naturelles ou synthétiques), physiologiquement acceptables et non toxiques, par l'intermédiaire de groupements réactifs complémentaires respectivement portés par la molécule porteuse et le peptide. Des exemples de groupements appropriés sont illustrés dans ce qui suit :

A titre d'exemple de molécules porteuses ou supports macromoléculaires entrant dans la constitution des conjugués selon l'invention, on mentionnera des protéines naturelles, telles que l'anatoxine tétanique, l'ovalbumine, des sérums albumines, des hémocyanines, etc...

A titre de support macromoléculaires synthétiques, on mentionnera par exemple des polylysines ou des poly(D-L-alanine)-poly(L-lysine).

La littérature mentionne d'autres types de supports macromoléculaires susceptibles d'être utilisés, lesquels présentent en général un poids moléculaire supérieur à 20 000.

Pour synthétiser les conjugués selon l'invention, on peut avoir recours à des procédés connus en soi, tels que celui décrit par FRANTZ et ROBERTSON dans Infect. and Immunity, 33, 193-198 (1981), ou celui décrit dans Applied and Environmental Microbiology, (octobre 1981), vol. 42, n° 4, 611-614 par P.E. KAUFFMAN en utilisant le peptide et la molécule porteuse appropriée.

Dans la pratique, on utilisera avantageusement comme agent de couplage les composés suivants, cités à titre non limitatif : aldéhyde glutarique, chloroformate d'éthyle, carbodiimides hydrosolubles [N-éthyl-N'(3-diméthylaminopropyl) carbodiimide, HCl], diisocyanates, bis-diazobenzidine, di- et trichloro-s-triazines, bromures de cyanogène, ainsi que les agents de couplage mentionnés dans Scand. J. Immunol., (1978), vol. 8, p. 7-23 (AVRAMEAS, TERNYNCK, GUESDON).

On peut avoir recours à tout procédé de couplage faisant intervenir d'une part une ou plusieurs fonctions réactives du peptide et d'autre part, une ou plusieurs fonctions réactives de molécules supports. Avantagusement, il s'agit des fonctions carboxyle et amine, lesquelles peuvent donner lieu à une réaction de couplage en présence d'un agent de couplage du genre de ceux utilisés dans la synthèse des protéines, par exemple, le 1-éthyl-3-(3-diméthylaminopropyl) carbodiimide, le N-hydroxybenzotriazole, etc... On peut encore avoir recours à la glutaraldéhyde, notamment lorsqu'il s'agit de relier entre eux des groupes aminés respectivement portés par le peptide et la molécule support.

Les peptides selon l'invention possèdent des propriétés antigéniques. Ils peuvent donc être utilisés dans des procédés de diagnostic pour la détection d'une infection par le virus HIV-2.

Comme on l'a déjà mentionné, des études ont permis de distinguer deux groupes de peptides pouvant être mis en oeuvre dans des procédés de détection d'anticorps contre le virus HIV-2 dans un fluide biologique humain, notamment un sérum ou un liquide céphalo-rachidien.

Un premier groupe (I) comprend les peptides gag1. Ces peptides reconnaissent des anticorps anti-HIV-2 et sont

donc capables de détecter une infection par HIV-2. Ils reconnaissent également dans une certaine mesure des anticorps anti-HIV-1.

Un second groupe (II) comprend des peptides qui correspondent plus particulièrement à ceux qui sont situés dans la partie transmembranaire et dans la fin de la partie externe de la protéine d'enveloppe. Ces peptides sont ceux précédemment désignés par env1, env2 et env3. Ils permettent la reconnaissance spécifique de la présence d'anticorps contre HIV-2 et permettent donc de discriminer chez une personne les infections passées ou présentes dues à un HIV, plus particulièrement entre celles qui ont été provoquées par un HIV-2 et celles qui l'ont été par un HIV-1.

L'invention concerne également une composition contenant au moins l'un des susdits peptides ou au moins un oligomère de ce peptide, caractérisée en ce qu'elle a la capacité d'être reconnue par des sérums d'origine humaine contenant des anticorps contre le virus HIV-2.

L'invention concerne un procédé de diagnostic in vitro un ou des peptides selon l'invention pour la détection d'anticorps contre HIV-2 dans des fluides biologiques, en particulier dans des sérums humains.

D'une façon générale le procédé de diagnostic in vitro ci-dessus comprend les étapes suivantes :

- la mise en contact de ce liquide biologique avec lesdits peptides,
- la détection de la présence éventuelle d'un complexe peptide-anticorps par des méthodes physiques ou chimiques, dans ledit liquide biologique.

Dans un mode de réalisation préféré de l'invention, la détection du complexe antigène-anticorps est réalisée grâce à des tests immunoenzymatiques (du type ELISA), immunofluorescents (du type IFA), radioimmunologiques (du type RIA) ou des tests de radioimmunoprécipitation (du type RIPA).

Ainsi l'invention concerne également tout peptide selon l'invention marqué à l'aide d'un marqueur adéquat du type enzymatique, fluorescent, radioactif, etc...

De telles méthodes comprennent par exemple les étapes suivantes :

- dépôt de quantités déterminées d'une composition peptidique selon l'invention dans les puits d'une microplaque de titration,
- introduction dans lesdits puits de dilutions croissantes du sérum devant être diagnostiqué,
- incubation de la microplaque,
- rinçages répétés de la microplaque,
- introduction dans les puits de la microplaque d'anticorps marqués contre des immunoglobulines du sang, le marquage de ces anticorps ayant été réalisé à l'aide d'une enzyme sélectionnée parmi celles qui sont capables d'hydrolyser un substrat en modifiant l'absorption des radiations de ce dernier, au moins à une longueur d'onde déterminée,
- détection, en comparaison avec un témoin de contrôle, de la quantité de substrat hydrolysé.

L'invention concerne également des coffrets ou kits pour le diagnostic in vitro de la présence d'anticorps contre les virus HIV-2 et, dans certains cas, HIV-1 dans un milieu biologique qui comprennent ;

- une composition peptidique selon l'invention,
- les réactifs pour la constitution du milieu propice à la réalisation de la réaction immunologique,
- les réactifs permettant la détection du complexe antigènes-anticorps produit par la réaction immunologique. De tels réactifs peuvent également porter un marqueur, ou être susceptibles d'être reconnus à leur tour par un réactif marqué. Plus particulièrement dans le cas où la composition polypeptidique sus-mentionnée n'est pas marquée.
- un tissu fluide biologique de référence dépourvu d'anticorps reconnus par la composition polypeptidique sus-mentionnée,

L'invention concerne les anticorps eux-mêmes formés contre les peptides de l'invention.

Il va de soi que cette production n'est pas limitée aux anticorps polyclonaux.

Elle s'applique encore à tout anticorps monoclonal produit par tout hybridome susceptible d'être formé, par des méthodes classiques, à partir des cellules spléniques d'un animal, notamment de souris ou de rat, immunisés contre l'un des peptides de l'invention, d'une part et des cellules d'une lignée de cellule myélome approprié d'autre part, et d'être sélectionné, par sa capacité à produire des anticorps monoclonaux reconnaissant le peptide initialement mis en œuvre pour l'immunisation des animaux.

L'invention concerne également des compositions immunogènes pour la production de vaccins dont le principe actif est constitué par au moins un peptide selon l'invention, ou un oligomère de ce peptide, ou un peptide sous forme conjuguée avec une molécule porteuse, caractérisées en ce qu'elles induisent la production d'anticorps contre les susdits peptides en quantité suffisante pour aussi inhiber les protéines du rétrovirus HIV-2, voire même le rétrovirus

## EP 0 750 041 A2

HIV-2 entrant en association avec un véhicule pharmaceutiquement acceptable.

Les compositions immunogènes pour la production de vaccins comprennent de façon avantageuse plus particulièrement au moins l'un des peptides précédemment désignés par env4, env5, env6, env7, env8, env9, env10, env11 voir des mélanges de ceux-ci.

5 Parmi ces peptides aptes à constituer des principes actifs de vaccins certains sont particulièrement préférés car ils possèdent une structure de base en acides aminés correspondant à des régions des glycoprotéines d'enveloppe qui présentent un important degré de conservation, non seulement dans les HIV-2, et dans les SIV, mais également dans les HIV-1. Ces peptides particulièrement préférés sont les peptides désignés par env4, certains peptides env5, env6 et env10.

10 Dans un mode de réalisation préféré de l'invention les peptides immunogènes (ou fragments de ces peptides) aptes à constituer des principes actifs de vaccins sont choisis parmi ceux dont les formules correspondent à des séquences qui, dans les glycoprotéines d'enveloppe de HIV-2, SIV et HIV-1 présentant une homologie en acides aminés supérieure à 50%, qui appartiennent à la partie externe de l'enveloppe du virus, qui sont dépourvus ou presque de délétions, et qui renferment des résidus de cystéine favorables à la stabilisation des liaisons et à la constitution de

15 boucles d'ancrage.

Les peptides suivants appartiennent à cette catégorie de peptides préférés.

### env4

20 XVTV-YGVP-W--ATZ

### env5

25 XL-NVTE-FZ

### env6

30 XKPCVKL-PLC-Z

### env7

35 XN-S-I-Z

### env10

40 XNC-GEF-YC-Z

### env11

45 XC-I-Q-IZ

Des compositions pharmaceutiques avantageuses sont constituées par des solutions, suspensions ou liposomes injectables contenant une dose efficace d'au moins un produit selon l'invention. De préférence, ces solutions, suspensions ou liposomes sont réalisés dans une phase aqueuse stérilisée isotonique, de préférence saline ou glucosée.

50 L'invention concerne plus particulièrement de telles suspensions, solutions ou forme liposome qui sont aptes à être administrées par injections intradermiques, intramusculaires ou sous-cutanées, ou encore par scarifications.

Elle concerne également des compositions pharmaceutiques administrables par d'autres voies, notamment par voie orale.

Les compositions pharmaceutiques selon l'invention, utilisables en tant que vaccins pour être efficaces dans la production d'anticorps contre le virus HIV-2, peuvent à titre d'exemple être administrées à des doses situées entre 10

55 et 500 µg/kg, de peptides selon l'invention, de préférence de 50 à 100 µg/kg.

Ces doses sont citées à titre d'exemple et ne possèdent en aucun cas un caractère limitatif.

Comme on l'a déjà indiqué plus haut les différents peptides qui ont été définis peuvent comprendre des modifications qui n'ont pas pour effet de modifier de façon fondamentale leurs propriétés immunologiques. Les peptides équi-

## EP 0 750 041 A2

valents qui en résultent entrent dans le champ des revendications qui suivent. A titre d'exemples de peptides équivalents on mentionnera ceux dont les structures en correspondance avec des régions des ADNc d'autres variants de HIV-2 de SIV ou de HIV-1, lorsque ces régions ont été mises en alignement dans des conditions semblables à celles qui ont été évoquées ci-dessus, à propos de HIV-2 ROD, SIV et HIV-1 BRU. A titre d'autres de ces peptides, on

5 mentionnera ceux dont les structures sont en correspondance avec de telles régions dans les ADNc qui ont fait l'objet de dépôts à la CNCM, notamment sous les numéros I-502, I-642 (HIV-2 IRMO), I-643 (HIV-2 EHO) ainsi que, dans les cas appropriés, des variants de HIV-1 qui ont fait l'objet de dépôts à la CNCM sous les numéros I-232, I-240, I-241, I-550, I-551.

Les peptides selon l'invention peuvent encore être définis par les formules suivantes (dans lesquels X, Z et les tirets "-" ont les significations sus-indiquées) :

10

15

20

25

30

35

40

45

50

55

EP 0 750 041 A2

XRV-AIEKYL-DQA-LN-WGCAFRQVCZ  
 XAIEKYL-DZ  
 5  
 X-LE-AQIQQEKNNMYELQKLSWZ  
 XQIQQEKNZ  
 10  
 XELGDYKLVEITPIG-APT--KR-----Z  
 XYKLVEITPIG-APT--KRZ  
 15  
 X----VTV-YGVP-W--AT--LFCA-Z  
 XVTV-YGVP-W--ATZ  
 20  
 X----E--L-NVTE-F--W-NZ  
 XL-NVTE-FZ  
 25  
 XL---S-KPCVKL-PLC-----Z  
 XKPCVKL-PLC-Z  
 XS-KPCVKL-PLC-Z  
 30  
 X---N-S-I---C-Z  
 XN-S-I-Z  
 35  
 XYC-P-G-A-L-C-N-TZ  
 X-----A-C-----W--Z  
 40  
 NKRPRQAWCWFKG-KWKD  
 45  
 X-G-DPE-----NC-GEF-YC-----NZ  
 X-----C-I-Q-I-----G---YZ

50

L'invention concerne également outre les peptides de SIV déjà décrits, les protéines codées par l'ADNc du virus SIV. Elle concerne également les protéines de tout virus immunologiquement étroitement apparenté à SIV-1mac, en particulier tout virus dont les protéines et les glycoprotéines d'enveloppe croisent immunologiquement et dont les ADNc présentent un pourcentage d'homologie d'au moins 95% et de préférence d'au moins 98%.

55

En particulier l'invention concerne :

- 1/ les protéines et glycoprotéines de l'enveloppe codées par le gène env et représentées à la figure 3,
- 2/ la protéine GAG représentée à la figure 4,

## EP 0 750 041 A2

- 3/ la protéine POL représentée à la figure 5,  
 4/ la protéine Q représentée à la figure 6,  
 5/ la protéine R représentée à la figure 7,  
 6/ la protéine X représentée à la figure 8,  
 7/ la protéine F représentée à la figure 9,  
 8/ la protéine TAT représentée à la figure 10,

Les acides aminés des protéines précitées de SIV, ont été représentées en alignement avec les séquences d'acides aminés des protéines correspondantes du virus HIV-2 les points verticaux figurant entre les deux séquences correspondent aux acides aminés communs entre les protéines des deux virus.

Les séquences d'ADNc codant pour les protéines précitées apparaissent sur la figure 1B. L'invention concerne, outre les séquences nucléiques précitées toute séquence nucléiques modifiée, qui code également pour les protéines du rétrovirus SIV ou d'un variant.

Ces séquences d'ADNc repérées par la numérotation figurant sur les séquences décrites précédemment (figure 1B) sont les suivantes :

-séquence codant pour <u>GAG</u> , nucléotides 551 à 2068					
20	-	"	"	<u>POL</u> ,	" 1726 à 4893
	-	"	"	<u>Q</u> ,	" 4826 à 5467
	-	"	"	<u>X</u> ,	" 5298 à 5633
	-	"	"	<u>R</u> ,	" 5637 à 5939
25	-	"	"	<u>F</u> ,	" 8569 à 9354
	-	"	"	<u>TAT-1</u>	" 5788 à 6084
	-	"	"	<u>ART-1</u>	" 6014 à 6130
30	-	"	"	<u>TAT-2</u>	" 8296 à 8391
	-	"	"	<u>ART-2</u>	" 8294 à 8548
	-	"	"	<u>ENV</u>	" 6090 à 8732

L'invention concerne donc naturellement les protéines précédemment décrites, lorsqu'elles sont obtenues à partir du virus SIV ou lorsqu'elles sont préparées par une méthode de synthèse, notamment par l'une des méthodes déjà citées en rapport avec la synthèse des peptides de plus petite taille.

L'invention concerne également l'utilisation des protéines précédentes pour le diagnostic de la présence éventuelle d'anticorps dirigés contre les protéines de HIV-2, voire contre HIV-2 en entier, ou pour certaines d'entre elles l'utilisation aux fins de diagnostic d'une infection due à l'un des virus HIV. Ainsi le peptide GAG codé par le gène correspondant peut être utilisé pour repérer la présence éventuelle d'anticorps anti-HIV-1 ou anti-HIV-2. Les protéines ENV sont utilisées de préférence pour le diagnostic spécifique d'une infection due à HIV-2 ou un de ses variants, parfois pour le diagnostic d'une infection par HIV-2 ou HIV-1.

L'invention concerne donc également un procédé de diagnostic in vitro de détection d'anticorps contre HIV-2 et éventuellement contre HIV-1 dans des fluides biologiques et en particulier dans des sérums humains. De tels procédés applicables pour l'utilisation des protéines précédentes de SIV comme protéines de diagnostic, ont déjà été décrits dans la présente invention.

L'invention concerne aussi des coffrets ou "kits" pour le diagnostic in vitro de la présence d'anticorps le virus HIV-2 et dans certains cas contre HIV-1 dans un milieu biologique. De tels kits mettant en oeuvre les peptides précédents ont également été décrits dans la présente invention.

L'invention concerne également des compositions immunogènes pour la production de vaccins, dont le principe actif est constitué de façon avantageuse par au moins la partie de la protéine ENV du virus SIV, cette protéine pouvant être sous forme conjuguée avec une molécule porteuse. Ces compositions immunogènes induisent la production d'anticorps contre le susdit peptide en quantité suffisante pour inhiber les protéines du rétrovirus HIV-2, voire le rétrovirus HIV-2 lui-même.

Toutefois l'utilisation aux fins de diagnostic des protéines de SIV n'est en rien limitée à celle des seuls protéines ENV ou GAG. D'autres protéines parmi celles décrites peuvent être envisagées, pour préparer des compositions de

diagnostic voire de vaccin.

## Revendications

1. Séquence de nucléotides caractérisée en ce qu'elle répond à la séquence nucléotidique représentée à la figure 1B ou à la figure 1C, ou en ce qu'elle contient la séquence nucléotidique représentée à la figure 1B ou à la figure 1C, ou en ce qu'il s'agit d'une partie de la séquence représentée à la figure 1B ou à la figure 1C, ladite partie de séquence codant pour un peptide reconnu par des anticorps présents dans le sérum d'un patient infecté par un rétrovirus HIV-2 ou étant utilisable comme sonde pour la détection dans un échantillon biologique, de la présence d'un rétrovirus HIV-2.
2. Séquence de nucléotides selon la revendication 1, caractérisée en ce qu'elle comprend l'une des séquences suivantes identifiées dans la figure 1B ou dans la figure 1C.

GAG s'étendant entre les nucléotides	550 à 2068
POL	1726 à 4893
Q	4826 à 5467
X	5298 à 5633
R	5637 à 5939
F	8569 à 9354
TAT-1	5788 à 6084
ART-1	6014 à 6130
TAT-2	8296 à 8391
ART-2	8294 à 8548
LTR	8950 à 9468 et 1 à 316
ENV	6090 à 8732

3. Séquence de nucléotides selon la revendication 1 ou la revendication 2, caractérisée en ce qu'elle comprend l'une des séquences suivantes identifiées dans la séquence de la figure 1A:
  - séquence correspondant aux nucléotides 7850 à 7927
  - séquence correspondant aux nucléotides 8030 à 8095
  - séquence correspondant aux nucléotides 7601 à 7636
  - séquence correspondant aux nucléotides 6170 à 6247
  - séquence correspondant aux nucléotides 6294 à 6349
  - séquence correspondant aux nucléotides 6392 à 6445
  - séquence correspondant aux nucléotides 6724 à 6763
  - séquence correspondant aux nucléotides 6794 à 6838
  - séquence correspondant aux nucléotides 7112 à 7162
  - séquence correspondant aux nucléotides 7253 à 7336
  - séquence correspondant aux nucléotides 7358 à 7426
  - séquence correspondant aux nucléotides 1535 à 1597
4. Séquence de nucléotides caractérisée en ce qu'il s'agit d'une séquence selon la revendication 1 modifiée par dégénéréscence du code génétique.
5. Séquence de nucléotides selon la revendication 1, caractérisée en ce qu'il s'agit de la séquence contenue dans le plasmide pSIV-1.1 (CNCM I-658) ou dans le plasmide pSIV-1.2 (CNCM 1-659).
6. Séquence de nucléotides selon la revendication 1, caractérisée en ce qu'elle code pour un polypeptide choisi parmi
  - POL<sub>ROD</sub> ou POL<sub>MAC</sub> représenté à la figure 5
  - Q<sub>ROD</sub> ou Q<sub>MAC</sub> représentée à la figure 6
  - R<sub>ROD</sub> ou R<sub>MAC</sub> représentée à la figure 7



## EP 0 750 041 A2

$X_{\text{ROD}}$  ou  $X_{\text{MAC}}$  représentée à la figure 8  
 $F_{\text{ROD}}$  ou  $F_{\text{MAC}}$  représentée à la figure 9  
 $TAT_{\text{ROD}}$  ou  $TAT_{\text{MAC}}$  représentée à la figure 10  
 $ART_{\text{ROD}}$  ou  $ART_{\text{MAC}}$  représentée à la figure 11.

5

7. Séquence de nucléotides selon l'une quelconque des revendications 1 à 6, caractérisée en ce qu'elle est marquée.

8. Utilisation d'une séquence de nucléotides selon l'une quelconque des revendications 1 à 7 comme sonde pour la détection dans un échantillon biologique, d'une infection par un rétrovirus HIV-2.

10

9. Acide nucléique recombinant caractérisé en ce qu'il comprend une séquence de nucléotides selon l'une quelconque des revendications 1 à 6, insérée dans un acide nucléique provenant d'un vecteur.

10. Procédé de diagnostic de la présence ou non du virus HIV-2 ou d'un variant, dans des échantillons de sérums ou d'autres liquides ou tissus obtenus à partir de patients suspectés d'être porteurs du virus HIV-2 comprenant :

15

- au moins une étape d'hybridation conduite dans des conditions stringentes, par mise en contact de l'ADN de cellules de l'échantillon du patient suspect avec une sonde selon la revendication 7 sur une membrane appropriée,
- le lavage de ladite membrane avec une solution assurant la conservation de ces conditions stringentes de l'hybridation,
- la détection de la présence ou non du virus HIV-2 par une méthode d'immunodétection.

20

25

11. Procédé selon la revendication 10, caractérisé en ce que l'étape d'hybridation est conduite dans des conditions non stringentes et en ce que le lavage de ladite membrane est effectué avec une solution assurant la conservation des conditions non stringentes.

30

35

40

45

50

55

FIG. 1.A

HIV2.ROD

R  
 GTCCCTCTGCCGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCTCCAGCACTAGCAGG  
 TAGAGCCTGGCTGTTCCCTGCTAGACTCTCACCAGCACTTGGCCGGTGGTGGGCAGACGG  
 100  
 CCCCACGCTTGCTTGCTTAAAAACCTCTTAATAAAGCTGCCAGTTAGAAGCAAGTTAAGT  
 R US  
 GTCTGCTCCCATCTCTCCTAGTCGCCCGCTGGTCAITCGGTGTTACCTGAGTAACAAGA  
 200  
 CCCTGGTCTGTTAGGACCCCTTCTTGCTTTGGGAAACCGAGGCAGGAAAAATCCCTAGCAGG  
 300  
 TTGGCCCTGAAACAGGGACTTGAAGAAGACTGAGAAGTCTTGAACACGCTGAGTGAAG  
 GCAGTAAGGGCGGCAGGAACAAACACGACGGAGTGTCTCTAGAAAGGGCGGGCCGAGG  
 400  
 TACCAAACGCAGCGTGTGAGCGCGGAGGAGAAGACGCTCTCGGCTGAGGTAAGTACCTA  
 CACCAAAAACTGTAGCCGAAAGGGCTTGGTATCTCTAATTTACACAGCTAGAAGATTGTG  
 500  
 MetGlyAlaArgAsnSerValLeuArgGlyLysLysAlaAspGluLeuGluArgIle  
 CGAGATGGCGCGGAGAACTCCGTCTTGAGAGGAAAGGAGATGAATTAGAAAGAAT  
 600  
 ArgLeuArgProGlyGlyLysLysLysTyrArgLeuLysHisIleValTrpAlaAlaAsn  
 CAGGTTACGGCCCGCGGAAAGAAAAAGTACAGGCTAAAAATATTGTGTGGCAGCGAA  
 LysLeuAspArgPheGlyLeuAlaGluSerLeuLeuGluSerLysGluGlyCysGlnLys  
 TAAATTGGACAGATTCCGATTAGCAGAGACCTGTTGGAGTCAAAAGAGGGTTGTCT/AAA  
 700  
 IleLeuThrValLeuAspProMetValProThrGlySerGluAsnLeuLysSerLeuPhe  
 AATTCTTACAGTTTTAGATCCAAATGGTACCGACAGGTTGAGAAAATTTAAAAAGTCTTTT  
 AsnThrValCysValIleTrpCysIleHisAlaGluGluLysValLysAspThrGluGly  
 TAATACTGTCTGCGTCATTTGGTGCATACACGCAGAAAGAGAAAGATACTGAAGC  
 800  
 AlaLysGlnIleValArgArgHisLeuValAlaGluThrGlyThrAlaGluLysMetPro  
 AGCAAAACAAATAGTGGGAGACATCTAGTGGCAGAAACACGAACTGCAGACAAAATGCC

FIG. 1A

SerThrSerArgProThrAlaProSerSerGluLysGlyGlyAsnTyrProValGlnHis  
 AAGCACAAGTAGACCAACAGCACCATCTAGCGAGAAGGGAGGAAATTACCCACTGCAACA  
 ValGlyGlyAsnTyrThrHisIleProLeuSerProArgThrLeuAsnAlaTrpValLys  
 TGTAGCGGCAACTACACCCATATACCGCTCAGTCCCCCAACCCATAATGCCTGGGTAAA  
 1000  
 LeuValGluGluLysLysPheGlyAlaGluValValProGlyPheGlnAlaLeuSerGlu  
 ATTACTAGAGCAAAAAAAGTTCCGGGCAGAAGTAGTCCAGGATTTCAGGCACCTCTCAGA  
 GlyCysThrProTyrAspIleAsnGlnMetLeuAsnCysValGlyAspHisGlnAlaAla  
 AGGCTGCACGCCCTATGATATCAACCAAATGCTTAATTGTGTGGGGGACCATCAAGCAGC  
 1100  
 MetGlnIleIleArgGluIleIleAsnGluGluAlaAlaGluTrpAspValGlnHisPro  
 CATGCAGATAATCAGGAGATTATCAATCAGGAAGCAGCAGAAATGGGATGTCCAACATCC  
 1200  
 IleProGlyProLeuProAlaGlyGlnLeuArgGluProArgGlySerAspIleAlaGly  
 AATACCAGGCCCTTACCAGCGGGGCAGCTTAGAGAGCCAAGGGGATCTGACATAGCAGC  
 ThrThrSerThrValGluGluGlnIleGlnTrpMetPheArgProGlnAsnProValPro  
 CACAACAAGCAGTAGAAGAACAGATCCAGTGGATGTTTAGGCCACAAAATCCTGTACC  
 1300  
 ValGlyAsnIleTyrArgArgTrpIleGlnIleGlyLeuGlnLysCysValArgMetTyr  
 AGTAGGAAACATCTATAGAAGATGGATCCAGATAGGATTGCAGAAGTGTGTCTCAGGATGTA  
 AsnProThrAsnIleLeuAspIleLysGlnGlyProLysGluProPheGlnSerTyrVal  
 CAACCCGACCAACATCCTAGACATAAAACAGGGACCAAGGAGCCGTTCCAAAGCTATGT  
 1400  
 AspArgPheTyrLysSerLeuArgAlaGluGlnThrAspProAlaValLysAsnTrpMet  
 AGATAGATTCTACAAAAGCTTGAGGGCAGAACAAACAGATCCAGCAGTGAAGAATTGGAT  
 1500  
 ThrGlnThrLeuLeuValGlnAsnAlaAsnProAspCysLysLeuValLeuLysGlyLeu  
 GACCCAAACACTGCTAGTACAAAATGCCAACCCAGACTCTAAATTAGTGCTAAAAGGACT  
 GlyMetAsnProThrLeuGluGluMetLeuThrAlaCysGlnGlyValGlyGlyProGly  
 AGGGATGAACCCCTACCTTAGAAGAGATGCTGACCGCCTGTCAGGGGTTAGGTGGGCCAGG  
 1600  
 GlnLysAlaArgLeuMetAlaGluAlaLeuLysGluValIleGlyProAlaProIlePro  
 CCAGAAAGGTAGATTAATGGCAGAGGCCCTGAAAGAGGTATAGGACCTGCCCCCTATCCC  
 PheAlaAlaAlaGlnGlnArgLysAlaPheLysCysTrpAsnCysGlyLysGluGlyHis  
 ATTGGCAGCAGCCAGCAGAGAAAGGCATTTAAATGCTGCAACTCTGCAAAGGAAGCGCA  
 1700  
 SerAlaArgGlnCysArgAlaProArgArgGlnGlyCysTrpLysCysGlyLysProGly  
 CTCGGCAAGACAATGCCGAGCACCTAGAAGGCAGGGCTGCTGGAAGTGTGGTAAGCCAGC  
 1800  
 ThrGlyArgPhePheArgThrGlyProLeuGly  
 HisIleMetThrAsnCysProAspArgGlnAlaGlyPheLeuGlyLeuGlyProTrpGly  
 ACACATCATGACAAACTGCCAGATAGACAGGCAGGTTTTTTAGGACTGGGCCCTTGGGG  
 LysGluAlaProGlnLeuProArgGlyProSerSerAlaGlyAlaAspThrAsnSerThr  
 LysLysProArgAsnPheProValAlaGlnValProGlnGlyLeuThrProThrAlaPro  
 AAAGAAGCCCCGCAACTTCCCGTGGCCCAAGTTCCGCAGGGGCTGACACCAACAGCACC  
 1900  
 ProSerGlySerSerSerGlySerThrGlyGluIleTyrAlaAlaArgGluLysThrGlu  
 ProValAspProAlaValAspLeuLeuGluLysTyrMetGlnGlnGlyLysArgGlnArg  
 CCCAGTCGATCCAGCAGTGGATCTACTGCAGAAATATATGCAGCAAGGGAAAAGACAGAG  
 ArgAlaGluArgGluThrIleGlnGlySerAspArgGlyLeuThrAlaProArgAlaGly  
 GluGlnArgGluArgProTyrLysGluValThrGluAspLeuLeuHisLeuGluGlnGly  
 AGAGCAGAGAGAGAGACCATACAAGGAAGTGACAGAGGACTTACTGCACCTCGAGCAGGG  
 (fig. 1A-suite 1)

GlyAspThrIleGlnGlyAlaThrAsnArgGlyLeuAlaAlaProGlnPheSerLeuTrp  
 GluThrProTyrArgGluProProThrGluAspLeuLeuHisLeuAsnSerLeuPheGly  
 CGAGACACCATACAGGGAGCCACCAACAGAGCTTGCTGCACCTCAATTCTCTCTTTGG  
 2100  
 LysArgProValValThrAlaTyrIleGluGlyGlnProValGluValLeuLeuAspThr  
 LysAspGln  
 AAAAGACCAGTAGTCACAGCATACATTGAGGGTCAGCCAGTAGAAGTCTTGTAGACACA  
 GlyAlaAspAspSerIleValAlaGlyIleGluLeuGlyAsnAsnTyrSerProLysIle  
 GGGGCTGACGACTCAATAGTAGCAGGAATAGAGTTAGGGAACAATTATAGCCCAAAAATA  
 2200  
 ValGlyGlyIleGlyGlyPheIleAsnThrLysGluTyrLysAsnValGluIleGluVal  
 GTAGGGGGAATAGGGGATTCTATAAATACCAAGGAATATAAAATGTAGAAATAGAAGTT  
 LeuAsnLysLysValArgAlaThrIleMetThrGlyAspThrProIleAsnIlePheGly  
 CTAAATAAAAGGTACGGGCCACCATAATGACAGGCGACACCCCAATCAACATTTTGGC  
 2300  
 ArgAsnIleLeuThrAlaLeuGlyMetSerLeuAsnLeuProValAlaLysValGluPro  
 AGAAATATTCTGACAGCCTTAGGCATGTCATTAAATCTACCAGTCGCCAAAGTAGAGCCA  
 2400  
 IleLysIleMetLeuLysProGlyLysAspGlyProLysLeuArgGlnTrpProLeuThr  
 ATAAAAATAATGCTAAAGCCAGGGAAGATGGACCAAACTGAGACAATGCCCTTAACA  
 LysGluLysIleGluAlaLeuLysGluIleCysGluLysMetGluLysGluGlyGlnLeu  
 AAAGAAAAATAGAAGCACTAAAAGAAATCTGTGAAAAAATGGAAGAAAGGCCAGCTA  
 2500  
 GluGluAlaProProThrAsnProTyrAsnThrProThrPheAlaIleLysLysLysAsp  
 GAGGAAGCACCTCCAATAATCTTATAATACCCCCACATTTGCAATCAAGAAAAAGGAC  
 LysAsnLysTrpArgMetLeuIleAspPheArgGluLeuAsnLysValThrGlnAspPhe  
 AAAAACAAATGGAGGATGCTAATAGATTTTCAGAGAACTAAACAAGGTAAGTCAAGATTTTC  
 2600  
 ThrGluIleGlnLeuGlyIleProHisProAlaGlyLeuAlaLysLysArgArgIleThr  
 ACAGAAATTCAGTTAGGAATTCACACCCAGCAGGGTTGCCAAGAAGACAAGAATTACT  
 2700  
 ValLeuAspValGlyAspAlaTyrPheSerIleProLeuHisGluAspPheArgProTyr  
 GTACTAGCTAGGGGATGCTTACTTTTCCATACCACTACATGAGGACTTTAGACCATAT  
 ThrAlaPheThrLeuProSerValAsnAsnAlaGluProGlyLysArgTyrIleTyrLys  
 ACTGCCATTACTCTACCATCAGTGAACAATGCAGAACCGGAAAAAGATACATATATAAA  
 2800  
 ValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnHisThrMetArgGlnVal  
 GTCTTGGCACAGGGATGGAAGGGATCACCAGCAATTTTCAACACACAATGAGACAGGTA  
 LeuGluProPheArgLysAlaAsnLysAspValIleIleIleGlnTyrMetAspAspIle  
 TTAGAACCATTTCAGAAAAGCAACAAGGATGTCATTATCATTTCAGTACATGGATGATATC  
 2900  
 LeuIleAlaSerAspArgThrAspLeuGluHisAspArgValValLeuGlnLeuLysGlu  
 TTAATAGCTAGTGACAGGACAGATTTAGAACATGATAGGGTAGTCTGCAGCTCAAGGAA  
 3000  
 LeuLeuAsnGlyLeuGlyPheSerThrProAspGluLysPheGlnLysAspProProTyr  
 CTTCTAAATGGCCTAGGATTTTCTACCCAGATGAGAAGTTCCAAAAAGACCCTCCATAC  
 HisTrpMetGlyTyrGluLeuTrpProThrLysTrpLysLeuGlnLysIleGlnLeuPro  
 CACTGGATGGGCTATGAACATATGGCCAATAATGGAAGTTGCAGAAAAATACAGTTGCCC  
 3100  
 GlnLysGluIleTrpThrValAsnAspIleGlnLysLeuValGlyValLeuAsnTrpAla  
 CAAAAAGAAATATGGACAGTCAATGACATCCAGAAGCTACTGCGTCTCTAAATTGGCCA

(fig.1A-suite 2)

AlaGlnLeuTyrProGlyIleLysThrLysHisLeuCysArgLeuIleArgGlyLysMet  
 GCACAACTCTACCCAGGGATAAAGACCAAACACTTATCTAGGTTAATCAGAGGAAAAATG  
 3200  
 ThrLeuThrGluGluValGlnTrpThrGluLeuAlaGluAlaGluLeuGluGluAsnArg  
 ACACCTCAGAGAAGAACTACACTGGACAGAATTAGCAGAAGCAGAGCTAGAAGAAAAACAGA  
 3300  
 IleIleLeuSerGlnGluGlnGluGlyHisTyrTyrGlnGluGluLysGluLeuGluAla  
 ATTATCCTAAGCCAGGAACAAGAGGGACACTATTACCAAGAAAGAAAAAGAGCTAGAAGCA  
  
 ThrValGlnLysAspGluGluAsnGluTrpThrTyrLysIleHisGlnGluGluLysIle  
 AGAGTCCAAAAGGATCAAGAGAATGAGTGGACATATAAAATACACCAGGAAGAAAAAATT  
  
 LeuLysValGlyLysTyrAlaLysValLysAsnThrHisThrAsnGlyIleArgLeuLeu  
 CTAAGTACAGGAAAAATATGCAAGGCTGAAAAACACCCATACCAATGGAATCAGATTGTTA  
  
 AlaGlnValValGlnLysIleGlyLysGluAlaLeuValIleTrpGlyArgIleProLys  
 GCACAGGTAGTTTCAAGAAATAGGAAAAGAAGCACTAGTCATTTGGGGACGAATACCAAAA  
 3500  
 PheHisLeuProValGluArgGluIleTrpGluGlnTrpTrpAspAsnTyrTrpGlnVal  
 TTTACCTACCAGTAGACAGAGACAAATCTGGGAGCAGTGGTGGGATAACTACTGCCAAGTGC  
 3600  
 ThrTrpIleProAspTrpAspPheValSerThrProProLeuValArgLeuAlaPheAsn  
 ACATGGATCCAGACTGGGACTTCGTCTCTACCCACCACTGGTCAGGTTAGCGTTTAAC  
  
 LeuValGlyAspProIleProGlyAlaGluThrPheTyrThrAspGlySerCysAsnArg  
 CTGGTAGGGGATCCTATACAGGTGCAGAGACCTTCTACACACATGGATCCTGCAATAGC  
 3700  
 GlnSerLysGluGlyLysAlaGlyTyrValThrAspArgGlyLysAspLysValLysLys  
 CAATCAAAAAGAAGGAAAAGCAGGATATGTAACAGATAGAGGGAAAGACAAGCTAAAGAAA  
  
 LeuGluGlnThrThrAsnGlnGlnAlaGluLeuGluAlaPheAlaMetAlaLeuThrAsp  
 CTAGAGCAAACTACCAATCAGCAAGCAGAACTAGAAGCCTTTGGCATGGCACTAACAGAC  
 3800  
 SerGlyProLysValAsnIleIleValAspSerGlnTyrValMetGlyIleSerAlaSer  
 TCGGGTCCAAAAGTTAATATTATAGTAGACTCACAGTATGTAATGGGCATCAGTGCAAGC  
 3900  
 GlnProThrGluSerGluSerLysIleValAsnGlnIleIleGluGluMetIleLysLys  
 CAACCAACAGAGTCAGAAAGTAAATAGTCAACCAGATCATAGAAGAAATGATAAAAAAG  
 4000  
 GluAlaIleTyrValAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlnGluVal  
 GAAGCAATCTATGTTGCATGGGTCCCAGCCCACAAAGGCATAGGGGGAAACAGGAAGTA  
 4100  
 AspHisLeuValSerGlnGlyIleArgGlnValLeuPheLeuGluLysIleGluProAla  
 CATCATTTAGTCAGTCAGGGTATCAGACAAGTGTGTTCTGGAAAAATAGAGCCCGCT  
  
 GlnGluGluHisGluLysTyrHisSerAsnValLysGluLeuSerHisLysPheGlyIle  
 CAGGAAGAACATGAAAAATATCATAGCAATGTAAAAGAACTGTCTCATAAATTTGGAATA  
 4200  
 ProAsnLeuValAlaArgGlnIleValAsnSerCysAlaGlnCysGlnGlnLysGlyGlu  
 CCCAATTTAGTGGCAAGGCAAAATAGTAAACTCATGTGCCCAATGTCAACAGAAAGGGGAA  
 4300  
 AlaIleHisGlyGlnValAsnAlaGluLeuGlyThrTrpGlnMetAspCysThrHisLeu  
 GCTATACATGGGCAAGTAAATGCAGAACTAGGCACTTGGCAAAATGGACTGCACACATTTA  
  
 GluGlyLysIleIleIleValAlaValHisValAlaSerGlyPheIleGluAlaGluVal  
 GAAGGAAAGATCATTATAGTAGCAGTACATGTTGCAAGTGGATTATAGAAGCAGAAGTC  
 4400  
 IleProGlnGluSerGlyArgGlnThrAlaLeuPheLeuLeuLysLeuAlaSerArgTrp  
 ATCCACAGCAATCAGGAAGACAAACAGCACTCTTCTATTGAAACTGGCAAGTAGGTGG  
 (fig.1A-suite 3)

ProIleThrHisLeuHisThrAspAsnGlyAlaAsnPheThrSerGlnGluValLysMet  
 CCAATAACACACTTGCATACAGATAATGGTGCCAACTTCACTTCACAGGAGGTGAAGATG  
 4400  
 ValAlaTrpTrpIleGlyIleGluGlnSerPheGlyValProTyrAsnProGlnSerGln  
 GTAGCATGGTGGATAGGTATAGAACAATCCTTTGGAGTACCTTACAATCCACAGAGCCAA  
 4500  
 GlyValValGluAlaMetAsnHisHisLeuLysAsnGlnIleSerGluThrIleValLeu  
 GGAGTAGTAGAAGCAATGAATCACCATCTAAAAAACCAATAAGTGAACAATAGTACTA  
 MetAlaIleHisCysMetAsnPheLysArgArgGlyGlyIleGlyAspMetThrProSer  
 ATGGCAATTCATTGCATGAATTTTAAAGAACGGGGGGAATAGGGCATATGACTCCATCA  
 4600  
 GluArgLeuIleAsnMetIleThrThrGluGlnGluIleGlnPheLeuGlnAlaLysAsn  
 GAAAGATTAATCAATATGATCACCACAGAACAGAGATACAATTCCTCCAAGCCAAAAAT  
 SerLysLeuLysAspPheArgValTyrPheArgGluGlyArgAspGlnLeuTrpLysGly  
 TCAAAATTAAGATTTTCGGGTCTATTTTCAGAGAAGGCAGAGATCAGTTGTGAAAGCA  
 4700  
 ProGlyGluLeuLeuTrpLysGlyGluGlyAlaValLeuValLysValGlyThrAspIle  
 CCTGGGGAAGTACTGTGAAAGGAGAGAGGAGCAGTCCTAGTCAAGGTAGGAACAGACATA  
 4800  
 LysIleIleProArgArgLysAlaLysIleIleArgAspTyrGlyGlyArgGlnGluMet  
 MetGluGluAspLysArgTrp  
 AAAATAATACCAAGAAGGAAAGCCAAGATCATCAGAGACTATGGAGGAAGACAAGAGATG  
 AspSerGlySerHisLeuGluGlyAlaArgGluAspGlyGluMetAla  
 IleValValProThrTrpArgValProGlyArgMetGluLysTrpHisSerLeuValLys  
 GATAGTGGTTCCACCTCGAGGGTGCCAGGGAGGATGGAGAAATGGCATAGCCTTGTCAA  
 4900  
 TyrLeuLysTyrLysThrLysAspLeuGluLysValCysTyrValProHisHisLysVal  
 GTATCTAAAATACAAAACAAAGGATCTAGAAAAGGTGTGCTATGTTCCCCACCATAAGGT  
 GlyTrpAlaTrpTrpThrCysSerArgValIlePheProLeuLysGlyAsnSerHisLeu  
 GGGATGGGCATGGTGGACTTGCAGCAGGGTAATATTTCCATTAAAGGAAACAGTCATCT  
 5000  
 GluIleGlnAlaTyrTrpAsnLeuThrProGluLysGlyTrpLeuSerSerTyrSerVal  
 AGAGATACAGGCATATTGGAACCTTAACACCAGAAAAAGGATGGCTCTCTCTTATTCACT  
 5100  
 ArgIleThrTrpTyrThrGluLysPheTrpThrAspValThrProAspCysAlaAspVal  
 AAGAATAACTTGGTACACAGAAAAGTTCTGGACAGATGTTACCCAGACTGTGCAGATGT  
 LeuIleHisSerThrTyrPheProCysPheThrAlaGlyGluValArgArgAlaIleArg  
 CCTAATACATAGCACTTATTTCCCTTGCTTTACAGCAGGTGAAGTAAGAAGAGCCATCAG  
 5200  
 GlyGluLysLeuLeuSerCysCysAsnTyrProArgAlaHisArgAlaGlnValProSer  
 AGGGCAAAAGTTATTGTCTGCTGCAATTATCCCGAGCTCATAGAGCCAGGTACCGTC  
 LeuGlnPheLeuAlaLeuValValValGlnGlnAsnAspArgProGlnArgAspSerThr  
 MetThrAspProArgGluThrValPro  
 ACTTCAATTTCTGGCCTTAGTGGTAGTCCAACAAAATGACAGACCCAGAGAGACAGTAC  
 5300  
 ThrArgLysGlnArgArgArgAspTyrArgArgGlyLeuArgLeuAlaLysGlnAspSer  
 ProGlyAsnSerGlyGluGluThrIleGlyGluAlaPheAlaTrpLeuAsnArgThrVal  
 CACCAGGAACAGCGGCGAACAGACTATCGGACAGGCCTTCGCCTGGCTAAACAGCACAG  
 5400  
 ArgSerHisLysGlnArgSerSerGluSerProThrProArgThrTyrPheProGlyVal  
 GluAlaIleAsnArgGluAlaValAsnHisLeuProArgGluLeuIlePheGlnValTrp  
 TAGAAGCCATAAACAGAGAGCAGTGAATCACCTACCCGAGAACTTATTTTCCAGGTGT  
 (fig.1A-suite 4)

AlaGluValLeuGluIleLeuAla  
 GlnArgSerTrpArgTyrTrpHisAspGluGlnGlyMetSerGluSerTyrThrLysTyr  
 GGCAGAGGTCTGGAGATACTGGCATGATGAACAAGGCATGTCAGAAAGTTACACAAAGT  
 5500  
 ArgTyrLeuCysIleIleGlnLysAlaValTyrMetHisValArgLysGlyCysThrCys  
 ATAGATATTTGTGCATAATACAGAAAGCAGTGTACATGCATGTTAGGAAAGGGTGTACTT  
 LeuGlyArgGlyHisGlyProGlyGlyTrpArgProGlyProProProProProProPro  
 GCCTGGGGAGGGGACATGGGCCAGGAGGTGGAGACCAGGGCCTCCTCCTCCTCCCCCTC  
 5600  
 MetAlaGluAlaProThrGluLeuProProValAspGlyThrProLeu  
 GlyLeuVal\*\*\*  
 CAGGTCTGTTCTAATGGCTGAAGCACCAACAGAGCTCCCCCGGTGGATGGGACCCCACT  
 ArgGluProGlyAspGluTrpIleIleGluIleLeuArgGluIleLysGluGluAlaLeu  
 GAGGGAGCCAGGGGATGAGTGGATAATAGAAATCTTGAGAGAAATAAAGAAGAAGCTTT  
 LysHisPheAspProArgLeuLeuIleAlaLeuGlyLysTyrIleTyrThrArgHisGly  
 MetGlu  
 AAAGCATTTTGACCCTCGCTTGCTAATTGCTCTTGGCAAATATATCTATACTAGACATGG  
 5800  
 AspThrLeuGluGlyAlaArgGluLeuIleLysValLeuGlnArgAlaLeuPheThrHis  
 ThrProLeuLysAlaProGluSerSerLeuLysSerCysAsnGluProPheSerArgThr  
 AGACACCCCTGAAGGCCGAGAGAGCTCATTAAAGTCCTGCAACGAGCCCTTTTCACCCA  
 PheArgAlaGlyCysGlyHisSerArgIleGlyGlnThrArgGlyGlyAsnProLeuSer  
 SerGluGlnAspValAlaThrGlnGluLeuAlaArgGlnGlyGluGluIleLeuSerGln  
 CTTGAGAGCAGCATCTGGCCACTCAAGAATTGGCCAGACAAGGGGAGGAAATCCTCTCTC  
 5900  
 AlaIleProThrProArgAsnMetGln  
 LeuTyrArgProLeuGluThrCysAsnAsnSerCysTyrCysLysArgCysCysTyrHis  
 AGCTATACCGACCCCTAGAAACATGCAATAACTCATGCTATTGTAAGCGATGCTGCTACC  
 6000  
 MetAsnGluArgAlaAsp  
 CysGlnMetCysPheLeuAsnLysGlyLeuGlyIleCysTyrGluArgLysGlyArgArg  
 ATTGTCAGATGTGTTTTCTAAACAAGGGGCTCGGGATATGTTATGAACCAAAGGGCAGAC  
 GluGluGlyLeuGlnArgLysLeuArgLeuIleArgLeuLeuHisGlnThrSerGluTyr  
 Met  
 ArgArgThrProLysLysThrLysThrHisProSerProThrProAspLys  
 GAAGAAGCACTCCAAAGAAAACCTAAGACTCATCCGTCTCCTACACCAGACAAGTGAGTAT  
 6100  
 AspGluSerAlaAlaTyrCysHisPheIleSer  
 MetAsnGlnLeuLeuIleAlaIleLeuLeuAlaSerAlaCysLeuValTyrCysThrGln  
 GATCAATCAGCTGCTTATTGCCATTTTATTAGCTAGTCTTGCTTAGTATATTGCACCCA  
 TyrValThrValPheTyrGlyValProThrTrpLysAsnAlaThrIleProLeuPheCys  
 ATATGTAAGTCTTTTCTATGGCGTACCCAGCTGGAAAAATGCAACCATTCGCCCTCTTTTG  
 6200  
 AlaThrArgAsnArgAspThrTrpGlyThrIleGlnCysLeuProAspAsnAspAspTyr  
 TGCAACCAGAAATAGGGATACTTGGGGAACCATAACAGTCTTGCCTGACAATGATGATTA  
 6300  
 GlnGluIleThrLeuAsnValThrGluAlaPheAspAlaTrpAsnAsnThrValThrGlu  
 TCAGGAAATAACTTTGAATGTAACAGAGGCTTTTGATGCATGGAATAATACAGTAACAGA  
 GlnAlaIleGluAspValTrpHisLeuPheGluThrSerIleLysProCysValLysLeu  
 ACAAGCAATAGAAGATGTCTGGCATCTATTGAGACATCAATAAAACCATCTGTCAAACCT  
 6400  
 (fig.1A-sufte 5)

ThrProLeuCysValAlaMetLysCysSerSerThrGluSerSerThrGlyAsnAsnThr  
 AACACCTTTATGTGTAGCAATGAAATGCAGCAGCACAGAGAGCAGCACAGGGGAACAACAC  
 ThrSerLysSerThrSerThrThrThrThrThrProThrAspGlnGluGlnGluIleSer  
 AACCTCAAAGAGCACAAGCACAACCACAACCACACCCACAGACCAGGAGCAAGAGATAAG  
 6500  
 GluAspThrProCysAlaArgAlaAspAsnCysSerGlyLeuGlyGluGluGluThrIle  
 TGAGGATACTCCATGCCGACGCCGAGACAACCTGCTCAGGATTGGGAGAGGAAGAAACGAT  
 6600  
 AsnCysGlnPheAsnMetThrGlyLeuGluArgAspLysLysLysGlnTyrAsnGluThr  
 CAATTGCCAGTTCAATATGACAGGATTAGAAAGAGATAAGAAAAACAGTATAATGAAAC  
 TrpTyrSerLysAspValValCysGluThrAsnAsnSerThrAsnGlnThrGlnCysTyr  
 ATGGTACTCAAAAGATGTGGTTTGTGAGACAAATAATAGCACAAATCAGACCCACTGTTA  
 6700  
 MetAsnHisCysAsnThrSerValIleThrGluSerCysAspLysHisTyrTrpAspAla  
 CATGAACCATTCGAACACATCAGTCATCACAGAATCACTGTGACAAGCACTATTGGGATGC  
 IleArgPheArgTyrCysAlaProProGlyTyrAlaLeuLeuArgCysAsnAspThrAsn  
 TATAAGCTTTAGATACTGTGCCACCCCGGTTATGCCCTATTAAAGATGTAATGATACCAA  
 6800  
 TyrSerGlyPheAlaProAsnCysSerLysValValAlaSerThrCysThrArgMetMet  
 TTATTTCAGGCTTTGCCACCAACTGTTCTAAAGTAGTAGCTTCTACATGCCACAGGATGAT  
 6900  
 GluThrGlnThrSerThrTrpPheGlyPheAsnGlyThrArgAlaGluAsnArgThrTyr  
 GGAAACGCAAACTTCCACATGGTTTGGCTTTAATGGCACTAGAGCAGAGAATAGAACATA  
 IleTyrTrpHisGlyArgAspAsnArgThrIleIleSerLeuAsnLysTyrTyrAsnLeu  
 TATCTATTGGCATGGCAGAGATAATAGAACTATCATCAGCTTAAACAAATATTATAATCT  
 7000  
 SerLeuHisCysLysArgProGlyAsnLysThrValLysGlnIleMetLeuMetSerGly  
 CAGTTTGCATTGTAAAGAGGCCAGGGAATAAGACACTGAAACAAATAATGCTTATGTCAGG  
 HisValPheHisSerHisTyrGlnProIleAsnLysArgProArgGlnAlaTrpCysTrp  
 ACATGTGTTTCACTCCCACTACCAGCCGATCAATAAAGAGCCAGACAAGCATGGTGCTG  
 7100  
 PheLysGlyLysTrpLysAspAlaMetGlnGluValLysGluThrLeuAlaLysHisPro  
 GTTCAAAGGCAAAATGGAAGACGCCATGCAGGAGGTGAAGGAAACCCCTTGCAAAACATCC  
 7200  
 ArgTyrArgGlyThrAsnAspThrArgAsnIleSerPheAlaAlaProGlyLysGlySer  
 CAGGTATAGAGCAACCAATGACACAAGGAATATTAGCTTTGCAGCGCCAGGAAAAGGCTC  
 AspProGluValAlaTyrMetTrpThrAsnCysArgGlyGluPheLeuTyrCysAsnMet  
 AGACCCAGAAGTAGCATACATGTGCTAACTGCAGAGCAGAGTTTCTCTACTGCAACAT  
 7300  
 ThrTrpPheLeuAsnTrpIleGluAsnLysThrHisArgAsnTyrAlaProCysHisIle  
 GACTTGGTTTCTCAATTGGATAGAGAATAAGACACACCCCAATTATGCACCGTGCCATAT  
 LysGlnIleIleAsnThrTrpHisLysValGlyArgAsnValTyrLeuProProArgGlu  
 AAAGCAAAATAATTAACACATGGCATAAGGTAGCGAGAAATGTATATTGCTCTCCAGGGA  
 7400  
 GlyGluLeuSerCysAsnSerThrValThrSerIleIleAlaAsnIleAspTrpGlnAsn  
 ACGCGAGCTGCTCTGCAACTCAACAGTAACCAGCATAATTGCTAACATTGACTGGCAAAA  
 7500  
 AsnAsnGlnThrAsnIleThrPheSerAlaGluValAlaGluLeuTyrArgLeuGluLeu  
 CAATAATCAGACAAACATTACCTTTAGTGCAGAGGTGGCAGAACTATACAGATTGGAGTT  
 GlyAspTyrLysLeuValGluIleThrProIleGlyPheAlaProThrLysGluLysArg  
 GGGAGATTATAAATTGGTAGAAATAACACCAATTGGCTTGCACCTACAAAAGAAAAAAG  
 7600

(fig.1A-suite 6)



TyrSerSerAlaHisGlyArgHisThrArgGlyValPheValLeuGlyPheLeuGlyPhe  
 ATACTCCTCTGCTCAGGGAGACATACAAGCGTGTGTTCTGCTAGGGTTCTTGGGTTT  
 LeuAlaThrAlaGlySerAlaMetGlyAlaAlaSerLeuThrValSerAlaGlnSerArg  
 TCTCGCAACAGCAGGTTCTCCAATGGCGCGCGCTCCCTGACCGTGTCTGGCTCAGTCCCG  
 7700  
 ThrLeuLeuAlaGlyIleValGlnGlnGlnGlnGlnLeuLeuAspValValLysArgGln  
 GACTTTACTGGCCGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGCTCAACAGACA  
 7800  
 GlnGluLeuLeuArgLeuThrValTrpGlyThrLysAsnLeuGlnAlaArgValThrAla  
 ACAAGAACTGTTGGGACTGACCGTCTCGGGAACGAAAAACCTCCAGGCAAGACTCACTGC  
 IleGluLysTyrLeuGlnAspGlnAlaArgLeuAsnSerTrpGlyCysAlaPheArgGln  
 TATAGAGAACTACCTACAGGACCAGGCGCGCTAAATTTCATGGGGATGTGGCTTTAGACA  
 7900  
 ValCysHisThrThrValProTrpValAsnAspSerLeuAlaProAspTrpAspAsnMet  
 AGTCTGCCACACTACTCTACCATGGGTAAATGATTTCCTTAGCACCTGACTGGGACAATAT  
 ThrTrpGlnGluTrpGluLysGlnValArgTyrLeuGluAlaAsnIleSerLysSerLeu  
 CACCTCGCAGGAATGGGAAAAACAAGTCCGCTAGCTGGAGGCAATATCAGTAAAAGTTT  
 8000  
 GluGlnAlaGlnIleGlnGlnGluLysAsnMetTyrGluLeuGlnLysLeuAsnSerTrp  
 AGAACAGGCACAAAATTCAGCAAGAGAAAAATATGTATGAAGTACAAAAATTAAATAGCTG  
 8100  
 AspIlePheGlyAsnTrpPheAspLeuThrSerTrpValLysTyrIleGlnTyrGlyVal  
 GGATATTTTGGCAATTGGTTTGACITTAACCTCCTGGGTCAAGTATATTCAATATGGAGT  
 LeuIleIleValAlaValIleAlaLeuArgIleValIleTyrValValGlnMetLeuSer  
 Val  
 GCTTATAATAGTAGCAGTAATAGCTTTAAGAATAGTGATATATGTAGTACAAATGTTAAG  
 8200  
 AlaCysPheLeuPheProProArgLeuTyrProThrAsp  
 ArgLeuArgLysGlyTyrArgProValPheSerSerProProGlyTyrIleGlnGlnIle  
 GlyLeuGluArgAlaIleGlyLeuPheSerLeuProProProValIleSerAsnArgSer  
 TAGGCTTAGAAAGGGCTATAGGCCTGTTTTCTCTTCCCCCCCCGGTTATATCCAACAGAT  
 ProTyrProGlnGlyProGlyThrAlaSerGlnArgArgAsnArgArgArgArgTrpLys  
 HisIleHisLysAspArgGlyGlnProAlaAsnGluGluThrGluGluAspGlyGlySer  
 IleSerThrArgThrGlyAspSerGlnProThrLysLysGlnLysLysThrValGluAla  
 CCATATCCACAAGGACCGGGGACAGCCAGCCAACGAAGAAACAGAAGAAGACGGTGAAG  
 8300  
 GlnArgTrpArgGlnIleLeuAlaLeuAlaAspSerIleTyrThrPheProAspProPro  
 AsnGlyGlyAspArgTyrTrpProTrpProIleAlaTyrIleHisPheLeuIleArgGln  
 ThrValGluThrAspThrGlyProGlyArg  
 CAACGGTGCAGACAGATACTGGCCCTGGCCGATAGCATATATACATTTCTGATCCGCCA  
 8400  
 AlaAspSerProLeuAspGlnThrIleGlnHisLeuGlnGlyLeuThrIleGlnGluLeu  
 LeuIleArgLeuLeuThrArgLeuTyrSerIleCysArgAspLeuLeuSerArgSerPhe  
 GCTGATTCGCCCTCTTGACCAGACTATACAGCATCTGCAGGCACTTACTATCCAGGAGCTT  
 ProAspProProThrHisLeuProGluSerGlnArgLeuAlaGluThr  
 LeuThrLeuGlnLeuIleTyrGlnAsnLeuArgAspTrpLeuArgLeuArgThrAlaPhe  
 CCTGACCCCTCCAACCTCATCTACCAGAATCTCAGAGACTGGCTGAGACTTAGAACAGCCTT  
 8500  
 LeuGlnTyrGlyCysGluTrpIleGlnGluAlaPheGlnAlaAlaAlaArgAlaThrArg  
 MetGlyAlaSerGlySerLysLysHisSerArgProProArgGlyLeuGlnGlu  
 CTTGCAATATGGGTGCGAGTGGATCCAAGAAGCATTCCAGGCCGCCCGGAGGGCTACAAG  
 (fig.1A-suite 7)

GluThrLeuAlaGlyAlaCysArgGlyLeuTrpArgValLeuGluArgIleGlyArgGly  
 ArgLeuLeuArgAlaArgAlaGlyAlaCysGlyGlyTyrTrpAsnGluSerGlyGlyGlu  
 AGAGACTCTTGGCGGCCGCTGCAGGGGCTTGTGGAGGGTATTGGAACGAATCGGGACGGG  
 8600  
 IleLeuAlaValProArgArgIleArgGlnGlyAlaGluIleAlaLeuLeu  
 TyrSerArgPheGlnGluGlySerAspArgGluGlnLysSerProSerCysGluGlyArg  
 AATACTCGCGGTTCGAAGAAGGATCAGACAGGGAGCAGAAATCGCCCTCCTGTGAGGGAC  
 8700  
 GlnTyrGlnGlnGlyAspPheMetAsnThrProTrpLysAspProAlaAlaGluArgGlu  
 GGCAGTATCAGCAGCGAGACTTTTATGAATACTCCATGGAAGGACCCAGCAGCAGAAAGGG  
 LysAsnLeuTyrArgGlnGlnAsnMetAspAspValAspSerAspAspAspGluVal  
 AGAAAAATTTGTACAGGCAACAAAATATCGATCATGTACATTGAGATCATGATGACCAAG  
 8800  
 ArgValSerValThrProLysValProLeuArgProMetThrHisArgLeuAlaIleAsp  
 TAAGAGTTTCTGTACACCAAAAGTACCCTAAGACCAATGACACATACATTGGCAATAG  
 MetSerHisLeuIleLysThrArgGlyGlyLeuGluGlyMetPheTyrSerGluArgArg  
 ATATGTCACATTTAATAAAAAACAAGGGGGGACTGGAAGGGATGTTTACAGTGAAAGAA  
 8900  
 HisLysIleLeuAsnIleTyrLeuGluLysGluGluGlyIleIleAlaAspTrpGlnAsn  
 GACATAAAATCTTAAATATATACTTAGAAAAAGGAAGAAGGGATAATTGCAGATTGGCAGA  
 9000  
 TyrThrHisGlyProGlyValArgTyrProMetPhePheGlyTrpLeuTrpLysLeuVal  
 ACTACACTCATGGGCCAGGAGTAAGATACCCAATGTTCTTTGGGTGGCTATGGAAGCTAG  
 ProValAspValProGlnGluGlyGluAspThrGluThrHisCysLeuValHisProAla  
 TACCAGTAGATGTCCACAAGAAGGGGAGGACACTGAGACTCACTGCTTAGTACATCCAG  
 9100  
 GlnThrSerLysPheAspAspProHisGlyGluThrLeuValTrpGluPheAspProLeu  
 CACAAACAAGCAAGTTTGATGACCCGCATGGGGAGACACTAGTCTGGGAGTTTGATCCCT  
 LeuAlaTyrSerTyrGluAlaPheIleArgTyrProGluGluPheGlyHisLysSerGly  
 TGCTGGCTTATAGTTACGAGGCTTTTATTCCGGTACCCAGAGGAATTTGGGCACAAGTCAG  
 9200  
 LeuProGluGluTrpLysAlaArgLeuLysAlaArgGlyIleProPheSer  
 GCCTGCCAGAGGAAGAGTGGAAAGGCGAGACTGAAAGCAAGAGGAATACCATTAGTTAAA  
 9300  
 GACAGGAACAGCTATACTTGGTCAGGGCAGGAAGTAACCTAACAGAAACAGCTGAGACTGC  
 AGGGACTTTCCAGAAAGGGGCTGTAACCAAGGGAGGCACATGGGAGGAGCTGGTGGGGAAC  
 9400  
 GCCCTCATATTCTCTGTATAAATATACCCGCTAGCTTGCATTGTACTTCGGTCGCTCTGC  
 GGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCTCCAGCAGTAGCAGGTAGAGCCTGG  
 9500  
 GTGTTCCCTGCTAGACTCTCACCAGCACTTGGCCGGTGCTGGCCAGACGGCCCCCAGCCTT  
 9600  
 GCTTGCTTAAAAACCTCCTTAATAAAGCTGCCAGTTAGAAGCA

(fig.1A-suite 8)

FIG 1B

ACTCGCTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCTCCAGCACTAGCAG  
 GTAGAGCCTGGGTGTTCCCTGCTAGACTCTCACCAGCACCTGGCCGGTGTGGGCAGAGT  
 GGCTCCACGCTTGGTTGCTTAAAGACCTCTCAATAAAGCTGCCATTTAGAACTAAGCTA  
 GTGTGTGTTCCATCTCTCTAGTCGCGGCTGGTCAACTCGGTACTCGGTAATAAAAAAG  
 ACCCTGGTCTGTTAGGACCTGGTCTGTTAGGACCTTTCTGCTTTGGGAAACCGAAGCA  
 GGAAAATCCCTAGCAGATTGGCGCCGAACAGGGACTTGAAGGAGAGTGAGAGACTCTG  
 AGTACGGCTGAGTGAAGGCAGTAAGGGCGGCAGGAACCAACCACGACGGAGTGCTCTAG  
 AAAGGCGCGGTGGTACACAGCGCGGTGAGGAGCGGGAGAGAAGAGGCTCTCTGGTTG  
 CAGGTAAGTGCAACACAAAAAGGAAATAGCTGTCTTTATCCAGGAAGGATAATAAGAT  
 GAGDMETGLYALAARGASHNSERVALLEUSERGLYLYSLYSALAAAPGLULEUGLU  
 AGAGTGGGAGATGGGCGCGAGAACTCCGTCTTGTGAGGAAGAAAGCAGATGAATTAGA  
 LYSILEARGLEUARGPROGLYGLYLYSLYSTYRMETLEULYSHISVALVALTRPALA  
 AAAAATTAGACTACGACCCGGCGGAAAGAAAAAGTACATGTTGAAGCATGTAGTATGGGC  
 ALAASNGLULEUASPARGPMEGLYLEUALAGLUSERLEULEUGLUASNLYSGLUGLYCYS  
 AGCAAATGAATTAGATAGATTGGATTAGCAGAAAGCCTGTTGGAGAACAAAGAAGGATG  
 GLNLYSILEUSERVALLLEUALAPROLEUVALPROTHRGLYSERGLUASNLEULYSSER  
 TCAAAAAATACTTTTGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAG  
 LEUTYRASNTHRYALCYSVALILETRPCYSILEHISALAGLUGLULYSVALLYSHISTHR  
 CCTTTATAACTGTCTGCTCATCTGGTGCATTACCGCAGAACAGAAAGTGAACACAC  
 GLUGLUALALYSGLNILEVALGLNARGHISLEUVALMETGLUTHRGLYTHRALAGLUTHR  
 TGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGATGGAACAGGAACAGCAGAAAC  
 METPROLYSTHRSERARGPROTHRALAPROPHE SERGLYARGGLYGLYASNTYRPROVAL  
 TATGCCAAAAACAAGTAGACCAACAGCACCATTAGCGGCAGAGGAGGAAATTACCCAGT  
 GLNGLNILEGLYGLYASNTYRTHRHISLEUPROLEUSERPROARGTHRLEUASHALATRP  
 ACAACAAATAGGTGGTAACTATACCCACCTACCATTAGCCCGAGAACATTAAATGCCTG  
 VALLYSLEULEGLUGLULYSLYSPHEGLYALAGLUVALVALSERGLYPHEGLNALALEU  
 GGTAATAATAGAGGAGAAGAAATTTGGAGCAGAAGTAGTGTCAGGATTTACGGCACT  
 SERGLUGLYCYSLEUPROTYRASPILEASNGLNMETLEUASNCYSVALGLYASPHISGLN  
 GTCAGAAAGGCTGCCTCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCA  
 ALAALAMETGLNILELEARGASPILEILEASNGLUGLUALAALASPTRPASPLEUGLN  
 AGCGGCTATGCAGATCATCAGAGATATTATAAATGAGGAGGCTGCAGATTGGGACTTGCA  
 HISPRDGLNGLNALAPROGLNGLNGLYGLNLEUARGGLUPROSERGLYSERASPILEALA  
 GCACCCACAACAAGCTCCACAACAAGGACAGCTTAGGGAGCCGTCAGGATCAGATATTGC  
 GLYTHRTHRSERTHRYALGLUGLUGLNILEGLNTRPMETTYRARGGLNGLNASHPROILE  
 AGGAACAACCTAGTACAGTAGAAGAACAATCCAGTGGATGTACAGACAACAGAACCCCAT  
 1300

FIG. 1A-

PROVALGLYASNIETRYRARGARTRP ILEGLNLEUGLYLEUGLNLYSCYSVALARGMET  
 ACCAGTAGCCAACATTTACAGGAGATGGATCCAACCTGGGGTTCACAAAATGTGTCAGAA  
 TYRASNPOTHRASNIETLEUASPVALLYSGLNGLYPROLYSGLUPROPHEGLNSETYR  
 GTATAACCCAAACAAACATTCTAGATGTAAACAAGGGCCAAAAGAGCCATTTACAGAGCTA  
 1400  
 VALASPARGPHEITYRLYSSERLEUARGALAGLUGLNTHRASPPROALAYALLYASNTRP  
 TGTAGACAGGTTCTACAAAAGTTAAGAGCAGAACAAAAGATCCAGCAGTAAAGAATTG  
 1500  
 METTHRGLNTHRLLEULEULEGLNASNALAASNPROASPCYSLSLEUYALLEULYSGLY  
 GATGACTCAAACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAGTGTGAAGG  
 LEUGLYTHRASNPOTHRLEUGLUGLUMETLEUTHRALACYSGLNGLYVALGLYGLYPRD  
 GCTGGGTACGAATCCCACCTAGAAGAAATGCTGACGGCCTGTCAAGGAGTAGGGGGGCC  
 1600  
 GLYGLNLYSALAARGLEUMETALAGLUALALEULYSGLUALALEUALAPROALAPROILE  
 AGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGGCCCTCGCACCAGCGCCAAT  
 POLVALLEUGLULEUTRP  
 PROPHEALAAALAGLNGLNLYSGLYPROARGLYSPROILELYSCYSTRPASNCYSGLY  
 CCCTTTTGACAGAGCCCAACAGAAGGGACCAAGAAAGCCAATTAAGTGTGGAATTGTGG  
 1700  
 GLUGLYARGTHRLLEUCYSLSALAMETGLNSERPRDLYSLYSTRGLYMETLEUGLUMET  
 LYSGLUGLYHISSEALAAARGGLNLYSARGALAPROARGAGGLNGLYCYSTRPLYSYCY  
 GAAGGAAGGACACTCTGCAAGGCAATGCAAGAGCCCAAGAAGACAGGATGCTGGAAATG  
 1800  
 TRPLYSASNGLYPROCYSTRGLYGLNMETPRDLYSGLNTHRGLYGLYPHEPHEARGPRO  
 GLYLYSMETASPHISYALMETALALYSCYSPROASNARGGLNALAGLYPHELEUGLYLEU  
 TGGAAAAATGGACCATGTTATGGCCAAATGCCCAACAGACAGCGGGTTTTTATGGCCT  
 TRPPROLEUGLYLYSGLUALAPROGLNPHÉPROHISGLYSESRERALASERGLYALAAAP  
 GLYPROTRPGLYLYSLYSPROARGASNPHÉPROMETALAGLNVALHISGLNGLYLEUTHR  
 TGGCCCTTGCGGAAGAAGCCCGCAATTTCCCATGGCTCAAGTGCATCAGGGGCTGAC  
 1900  
 ALAASNCYSERPROARGARGTHRSERCYSGLYSERALALYSGLULEUMISALALEUGLY  
 PROTHRALAPROGLUGLUPROALAVALASPLEULEULYSASNTYRMETHISLEUGLY  
 GCCAACTGCTCCCCAGAGAAGAACAGCTGTGATCTGCTAAAGAACTACATGCACTTGGG  
 GLNALAALAGLUARGLYSGLNARGGLUALALEUGLNGLYGLYASPARGGLYPHEALALA  
 LYSGLNGLNARGGLUSERARGGLYLYSPOTYRLYSGLUVALTHRGLUASPLEULEUMIS  
 CAAGCAGCAGAGAGAAAGCAGAGGGAAGCCTTACAAGGAGGTGACAGAGGATTGCTGCA  
 2000  
 PROGLNPHÉSERLEUTRPARGARGPROVALYALTHRALAHISILEGLUGLYGLNPROVAL  
 LEUASNSEERLEUPHEGLYGLYASPLN  
 CCTCAATTCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTA  
 2100  
 GLUVALLEULEUASPTHRLYALAAASPSPSERILEVALTHRGLYILEGLULEUGLYPRO  
 GAAGTATTATTAGATACAGGGGCTGATGATTCTATTGTAAACAGGAATAGAGTTAGGTCCA  
 HISTYRTHRPRDLYSILEVALGLYGLYILEGLYGLYPHEILEASNTHRLYSGLUTYRLYS  
 CATTATACCCCAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAA  
 2200  
 ASNYALGLUILEGLUVALLEUGLYLYSARGILELYSGLYTHRILEMETTHRGLYASPTH  
 AATGTAGAAATAGAAGTTTTAGGCAAAAGGATTAAAGGGACAATCATGACAGGGGACACC  
 PROILEASNIPEHEGLYARGASNLEULEUTHRALALEUGLYMETSERLEUASNLEUPRO  
 CCGATTAAATTTTTGGTAGAAATTTACTAACAGCTCTGGGGATGCTCTCTAAATCTTCCC  
 2300  
 ILEALALYSVALGLUPROVALLYSSEPROLEULYSPOGLYLYSASPGLYPROLYSLEU  
 ATAGCTAAGGTAGAGCCTGTAAAGTCGCCCTTAAAGCCAGGAAAGGATGGACAAAATTG  
 2400  
 LYSGLNTRPPROLEUSERLYSGLULYSILEVALAELEUARGGLUILECYSGULYSMET  
 AAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAAGAGAAATCTGTGAAAAGATG

(fig. 12-suite 1)

GLULYSASPGLYGLNLEUGLUGLUALAPROPROTHRASHNPROTYRASNTHRPROTHRPH  
 GAAAAAGATGCTCAGTTGGAGGAAGCTCCCCCGACCAATCCATATAACACCCCCACATT  
 2500  
 ALAILELYSLYSLYSASPLYASNLYSTRPARGMETLEUILEASPPHEARGGLULEUASN  
 GCTATAAAGAAAAAGGATAAAAACAAATGCAGAATGCTGATAGATTTTAGGGAACATAAT  
 ARGVALTHRGLNASPPHETHRGLUYALGLNLEUGLYILEPROMISPROALAGLYLEUALA  
 AGGGTCACTCAAGACTTTACGGAAGTCCAATTAGGAATACCACACCTGCAGGACTAGCA  
 2600  
 LYSARGLYSARGILETHRVALLUEA SPILEGLYASPALATYRPHESERILEPROLEUASP  
 AAAAGGAAAAAGGATTACAGTACTGGATATAGGTGACGCATATTTCTCTATACCTCTAGAT  
 2700  
 GLUGLUPHEARGGLNTRYRTHRALAPHETHRLEUPROSERVALASNASNALAGLUPROGLY  
 GAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGCCAGGA  
 LYSARGTYRILETYRLYSVALLEUPROGLNGLYTRPLYSGLYSERPRODALILEPHEGLN  
 AAACGATACATTTATAAGGTTCTGCCTCAGGCATGGAAGGGGTCACAGCCATCTTCCAA  
 2800  
 TYRTHRMETARGHISYALLEUGLUPROPHEARGLYSALAASNPROASPVALTHRLEUVAL  
 TACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAAAATCCAGATGTGACCTTAGTC  
 GLNTRYMETASPPSILELEUILEALASERASPARGTHRASPLEUGLUHISASPARGVAL  
 CAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACAGGGTA  
 2900  
 VALLEUGLNLLEULYSGLULEULEUASNSERILEGLYPHESERSERPROGLUGLULYSPE  
 GTTTTACAGTTAAAAGAACTCTTAAATAGCATAGGGTTTTTCATCCCCAGAGAGAAATTC  
 3000  
 GLNLYSASPPOPROPHEGLNTRPMETGLYTYRGLULEUTRPPROTHRLYSTRLYSLEU  
 CAAAAAGATCCCCCATTTCAATGGATGGGCTACGAATTGTGCCCGACAAATGGAAGTTG  
 GLNLYSILEGLULEUPROGLNARGGLUTHRTRPTHRVALASNASPILEGLNLYSLEUVAL  
 CAAAAGATAGAGTTGCCACAAAGAGAGACCTGGACAGTGAATGATATACAGAAGTTAGTA  
 3100  
 GLYVALLEUASNTRPALAALAGLNILETYRPROGLYILELYSTHRLYSHISLEUCYSARG  
 GGAGTATTAATTTGGGCAGCTCAAATTTATCCAGGTATAAAACCAACATCTCTGTAGG  
 LEUILEARGGLYLYSMETTHRLEUTHRGLUGLUALGLNTRPTHRGLUMETALAGLUALA  
 TTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCACTGGACTGAGATGGCAGAAGCA  
 3200  
 GLUTYRGLUGLUASNLYSILEILEUSERGLNGLUGLNLGLUGLYCYSTYRTRYRGLNGLU  
 GAATATGAGGAAAAATAAATAATTCTCAGTCAGGAACAAGAAGGATGTTATTACCAAGAA  
 3300  
 SERLYSPROLEUGLUALATHRYVALILELYSSERGLNASPASNGLNTRPSERTYRLYSILE  
 AGCAAGCCATTAGAAGCCACGGTGATAAAGAGTCAGGACAATCAGTGGTCTTATAAAAT  
 HISGLNGLUASPLYSILELEULYSVALGLYLYSPHEALALYSILELYSASNTHRHISTHR  
 CACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAGATAAAGAATACACATACC  
 3400  
 ASNGLYVALARGLEULEUALAHISVALILEGLNLYSILEGLYLYSGLUALAILEVALILE  
 AATGGAGTTAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAGTGATC  
 TRPGLYGLNVALPROLYSPHEHISLEUPROVALGLULYSASPVALTRPGLUGLNTTRP  
 TGGGGACAGCTCCCAAAATTCACCTACCAGTTGAGAAGGATGTATGGGAACAGTGGTGG  
 3500  
 THRASPTYRTRPGLNVALTHRTRPILEPROGLUTRPASPPHEILESERTHRPROPROLEU  
 ACAGACTATTGGCAGGTAACCTGGATACCGAATGGGATTTCTCTCAACACCACCATTA  
 3600  
 VALARGLEUVALPHEASNLEUVALLYSASPPROILEGLUGLYGLUGLUTHRTYRTRYVAL  
 GTAAGATTAGTCTCAATCTAGTGAAGGACCTATAGAGGGAGAGAAACCTATTATGTA  
 ASPGLYSERCYSSERLYSGLNLSERLYSGLUGLYLYSALAGLYTYRILETHRASPARGGLY  
 GATGCATCATGTAGTAAACAGTCAAAGAGGAAAAGCAGGATATATCACAGACAGGGGC

(fig. 1B-suite 2)

3700  
 LYSASPLYSVALLYSVALLEUGLUGLNTHRTHRASNGLNGLNALAGLULEUGLUALAPHE  
 AAAGACAAGGTAAAAGTGTACAACAGACTACTAATCAACAAGCAGAATTGGAAGCATT  
 LEUMETALALEUTHRASPSERGLYPROLYSALAASNILEILEVALASPSERGLNTRYVAL  
 CTCATGGCATTGACAGACTCAGGGCCAAGGCCAATATTATAGTAGACTCACAAATATGTT  
 3800  
 METGLYILEILETHRGLYCYSPROTHRGUSERGLUSERARGLEUVALASNGLNILEILE  
 ATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAACCAATAATA  
 3900  
 GLUGLUMETILELYSLYSTHRGLUILETYRYALALATRPVALPRODALAHISLYSGLYILE  
 GAAGAAATGATCAAAAAGACAGAAATTTATGTGGCATGGGTACCAGCACACAAGGTATA  
 GLYGLYASNGLNGLUILEASPHISLEUVALSERGLNGLYILEARGGGLNVALLEUPHELEU  
 GGAGGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTCTCTCTCTG  
 4000  
 GLULYSTILEGLUPRODALAGLUGLUMISSERLYSTYRHISSEASNILELYSGLULEU  
 GAAAGATAGAGCCAGCACAGAAGAACATAGTAAATACCATAGTAACATAAAGAATTG  
 VALPHELYSPHEGLYLEUPROARGLEUVALALALYSGLNILEVALASPTHRCYSASPLYS  
 GTATTCAAATTTGGATTACCCAGACTAGTGGCCAAACAGATAGTAGACACATGTGATAAA  
 4100  
 CYSHISGLNLYSGLYGLUALAILEMISGLYGLNVALASNSERASPLEUGLYTHRTRPGLN  
 TGTCTCAAAAAGGAGAAGCTATACATGGCCAGGTAAATTCAGACCTAGGGACTTGGCAA  
 4200  
 METASPCYSTHRMISLEUGLUGLYLYSILEVALILEVALALAVAHISVALALASERGLY  
 ATGGATTGTACCATCTAGAGGGAAAAATAGTCATAGTTGCAGTACATGTAGCTAGTGA  
 PHEILEGLUALAGLUVALILEPROGLNGLUTHRGLYARGGLNTHRALALEUPHELEULEU  
 TTCATAGAAGCAGAAGTAATTCACAAGAAACAGGAAGACAGACAGCACTATTTCTGTTA  
 4300  
 LYSLEUALASERARGTRPPROILETHRHISLEUMISTHRASPSNGLYALAAASNPHLEALA  
 AAATTGGCAAGCAGATGGCCTATTACACATCTGCACACAGATAATGGTGCTAACTTTGCT  
 SERGLNGLUVALLYSMETVALALATRPTRPALAGLYILEGLUHISTHRPHEGLYVALPRO  
 TCGCAAGAAGTAAAGATGGTGCATGGTGGGCAGGATAGAGCACACCTTTGGGGTACCA  
 4400  
 TYRASNPGLNSERGLNGLYVALVALGLUALAMETASNNHISHISLEULYSASNGLNILE  
 TACAATCCACAGAGTCAGGGAGTAGTGGGAAGCAATGAATCACCACCTCAAAAATCAAATA  
 4500  
 ASPARGILEARGGLUGLNALAASNSEVALGLUTHRILEVALLEUMETALAVAHISYCS  
 GATAGAATCAGGGAACAAGCAAATTCAGTAGAAACCATAGTATTAATGGCAGTTCATTGC  
 METASNPHELYSARGARGGLYGLYILEGLYASPHETTHRPRDALAGLUARGLEULEASN  
 ATGAATTTTAAAACAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAATAAC  
 4600  
 METILETHRTHRGLUGLUGLUILEGLNPHEGLNGLNLSERLYSASNLERLYSPHELYSASN  
 ATGATCACTACAGAACAAGAAATACAATTTCAACAATCAAAAACCTCAAAATTTAAAAAT  
 PHEARGVALTYRTYRARGGLUGLYARGASPGLNLEUTRPLYSGLYPROGLYGLULEULEU  
 YTTCCGGCTCTATTACAGAGAAGGCAGAGATCAGCTGTGGAAGGGACCCGGTGAGCTATTG  
 4700  
 TRPLYSGLYGLUGLYALVALILELEULYSVALGLYTHRASPILELYSVALVALPROARG  
 TGGAAAGGGGAAGGAGCAGTCATCTTAAAGGTAGGAACAGACATTAAGGTAGTACCAGG  
 4800  
 ARGLYSALALYSILEILELYSASPTYRGLYGLYGLYLYSGLUMETASPSERSERSEHIS  
 QMETGLUGLUGLULYSARGTRPILEVALVALPROTHR  
 AGAAAGGCTAAAATATCAAAGATTTATCGAGGAGGAAAAGAGATGGATAGTAGTTCCAC  
 METGLUASPTHRGLYGLUALAARGGLUVALALA  
 TRPARCILEPROGLUARGLEUGLUARGTRPHISSERLEUILELYSTYRLEULYSTYRYS  
 ATGGAGGATACCGGAGAGGCTAGAGAGGTGGCATAGCCTCATAAAATATTTGAAATATAA  
 4900

(fig.1B-suite 3)

THRLYSASPLEUGLNLYSALACYSTYRVALPROHISMISLYSVALGLYTRPALATRPTRP  
 AACTAAAGATCTACAAAAGGCTTGCTATGTGCCCCATCATAAGGTCCGATGGCCATGGTG  
 THRCYSSERARGVALILEPHEPROLEUGLNGLYSERHISLEUGLUVALGLNGLYTYR  
 GACCTGCAGCAGAGTAATCTTCCCCTACAGGAAGGAAGCCATTTAGAAGTACAAGGGTA  
 5000  
 YRPASNLEUTHRPROGLUARGGLYTRPLEUSERTHRTYRALAVALARGILETHRTPTYR  
 TTGGAATTTGACACCAGAAAGAGGGTGGCTCAGTACTTATGCAGTGAGGATAACCTGGTA  
 5100  
 SERLYSASPHEPTRPTHRASPVALTHRPROGLUTYRALAASPILELEULEUMISSERTHR  
 CTCAAAGGACTTTTGGACAGATGTAAACACCAGAATATGCAGATATTTACTGCATAGCAC  
 TYRPHROCYSPHETHRALAGLYGLUVALARGARGALAILEARGGLYGLUARGLEULEU  
 TTATTTCCCTTGCTTTACAGCGGGAGAGTGAAGGGCCATCAGGGGAGAACGACTGCT  
 5200  
 SERCYSCYARGPHEPROARGALAHISLYSHISGLNVALPROSERLEUGLNTYRLEUALA  
 GCTTGCTGCAGGTTCCCAAGAGCTCATAAGCACCAGGTACCAAGTCTACAGTACTTAGC  
 LEUARGVALVALSERHISVALARGSERGLNGLYGLUASNPROTHRTPLYSGLNTRPARC  
 XMETSERASPPROARGGLUARGILEPROPROGLYASNSEGLYGLU  
 ACTGAGAGTAGTAAGTCATGTGATGCCAGGGAGAGAATCCACCTGGAAACAGTGGAG  
 5300  
 ARGASPNARGARGSERLEUARGVALALALYSGLNASNSERARGGLYASPLYSGLNARG  
 GLUTHRIEGLYGLUALAPHEGLUTRPLEUASNARGTHRYVALGLUGLUILEASNARGGLU  
 AAGACACAATAGGAGAAGCCTTCGAGTCCCTAACAGAACAGTAGAGGAGATAAACAGAG  
 5400  
 GLYGLYLYSPROPROTHRGLUGLYALAAASNPHROGLYLEUALALYSVALLEUGLYILE  
 ALAYALASNHISLEUPROARGGLULEULEPHEGLNVALTRPGLNARGSERTRPGLUTYR  
 AGGCGGTAACACCTACCGAGGGAGCTAATTTCCAGGTTTGGCAAAGGCTCTGGGAAT  
 LEUALA  
 TRPHISASPLUGLNGLYMETSERGLNSERTYRTHRLYSTYRARGTYRLEUCYSLEULE  
 ACTGGCATCATGAACAAGGATGTCAAAAGCTATACAAAATACAGATACTTGTTTAA  
 5500  
 GLNLYSALALEUPHEMETHISCYSLYSGLYCYSGCYSLEUGLYGLUGLYHISGLY  
 TACAAAAGGCTTTATTTATGCTTGAAGAAAGGCTGTAGATGCTAGGGGAAGGACAG  
 ALAGLYGLYTRPARGPARGGLYPROPRDPROPRDPROPRDPROGLYLEUALA R METGLU  
 GGGCAGGGGATGGAGACCAGGACCTCCTCCTCCCCCTCAGGACTAGCATAAATGG  
 5600  
 GLUARGPROPROGLUASNGLYGLYPROGLNARGGLUPROTRPASPLUTRPAVALGLU  
 AAGAAAGACCTCCAGAAATGAAGGCCACAAAGGGAACCATGGGATGAGTGGGTAGTGG  
 5700  
 VALLEULYSGLULEULYSGLUGLUALALEULYSHISPHEASPPROARGLEULEUTHRALA  
 AAGTTCTGAAAGAACTGAAAGAAGAAGCTTTAAAGCATTGTGATCCTCGGCTTCTAACC  
 TAT1 METGLUTHRPRDLEUARGGLUGLNGLUASNSER  
 LEUGLYASNHISILETYRASNARGHISGLYASPTHREUGLUGLYALAGLYGLULEULE  
 CACTTGTAATCATATCTATAATAGACATGGAGACACCCTTGAGGGAGCAGGAGAACTCA  
 5800  
 LEUGLUSERSERASNGLUARGSERSEITYRILESERGLUALAALAALAILEPROGLU  
 ARGILELEUGLNARGALALEUPHEILEHISPHEARGSERGLYCYSSERHISSEARGILE  
 TTAGAATCCTCCAACGAGCGCTCTCATACATTTCAGAAGCGGCTGCAGCCATTCCAGAA  
 SERALAASNLEUGLYGLUGLUILELEUSERGLNLEUTYRARGPROLEUGLUACYSTYR  
 GLYGLNPROGLYGLYGLYASNPROLEUSERTHRIEPROPROSERARGSERMETLEU  
 TCGGCCAACCTGGGGGAGGAAATCCTCTCTCACTATACCGCCCTCTAGAAGCATGCTAT  
 5900  
 ASNTHRCYSTYRCYSLYSLSYSCYSTYRHISCYSGLNPHCYSPHELEULYSLSGLY  
 AACACATGCTATTGCAAAAAGTGTGCTACCATGGCAGTTTGTCTTTCTTAAAAAGGGC  
 6000  
 LEUGLYILESEITYRGLULYSSEHISARGARGARGARGTHRPRDLYSLSALALYSALA  
 ARTMETARGSERHISTHRGLYGLUGLUGLULEUARGARGGLEUARGLEU

(fig.1B-suite 4)

TTGGGGATAAGTTATGACAAGTCACACAGGAGAAGAAGAACTCCGAACAAGGCTAAGGCT  
 ASNTHRSERSEALASERASNGLU  
 ILEHISLFULEUMISGLNTHRSERLYSTYRGLYLEUSERTRPLYSSERALAALATYRARG  
 ENV METGLYCYSLEUGLYASNGLNLEULEULEALA  
 AATACATCTTCTGCATCAAAACGAGTAAGTATGGGTGCTTGGAAATCAGCTGCTTATCG  
 6100  
 HISLEULEU  
 ILECYSSERLYSCYSLEUTRPILEILECYSILEGLNTRYVALTHRVALPHETRYGLYVAL  
 CCATCTGCTCTAAGTGCTATGGATTATTTGTATTCAATATGTCACAGCTTTTATGGTG  
 PROALATRPARGASNALATHRILEPROLEUPHECYSALATHRLYSASNARGASPTHTRP  
 TACCAGCTTGGAGGAATGGACAATTCCCTCTTCTGTGCAACCAAGAATAGGGATACCT  
 6200  
 GLYTHRTHRGLNLCYSLEUPROASPASNASPASPTYSERGLULEUALALEUASNVALTHR  
 GGGGAACAACCTCAGTGCCTACCAGATAATGATGATTATTCAGAATTGGCCCTTAATGTTA  
 6300  
 GLUSERPHEASPALATRPGLUASNTHRVATHRGLUGLNALALEGLUASPYALTRPGLN  
 CAGAAAGCTTTCATGCTTGGGAGAATACAGTCACAGAACAGGCAATAGAGGACCTATGGC  
 LEUPHEGLUTHRSERILELYSPROCYSVALLYSLEUSERPROLEUCYSILETHRMETARG  
 AACTCTTTGAGACCTCAATAAAGCCTTGTGTAATAATATCCCATTAATGACTACTATGA  
 6400  
 CYSASNLYSSERGLUTHRASPLYSTRPGLYLEUTHRLYSERSERTHRTHRALASER  
 CATGCAATAAAGTGACACAGATAAATGGGGATTGACAAAATCATCAACAACAGCAT  
 THRTHRTHRTHRTHRALALYSSEVALGLUTHRARGASPILEVALASNGLUTHRSER  
 CAACAACAACAACAACAGCAAAATCAGTAGAGACAAGACATAGTCAATGAGACTA  
 6500  
 PROCYSVALVALHISASPASNLCYSYTHRGLYLEUGLUGLUPROMETILESERCYSLS  
 GTCTTGTGTAGTTCATGATAATTGCACAGGCTTGGAAACAGAGCAATGATAAGCTGTA  
 6600  
 PHEASNMETTHRGLYLEULYSARGASPLYSLYSLYSLUTYRASNGLUTHRTRPTYSER  
 AATCAACATGACAGGGTAAAAAGAGACAAGAAAAAGGAGTACAATGAAACTTGCTACT  
 ALAASPLEUVALCYSGLUGLNGLYASNSETHRGLYASNGLUSERARGCYSYRMETASN  
 CTGCAGATCTGGTTTGTGAACAAGGGAATAGCACTGGTAATGAAAGTAGATGTTACATGA  
 6700  
 HISCYASNTHRSERVALILEGLNGLUCYSCYSASPLYASPTYRTRPASPALAILEARG  
 ATCACTGAATACTTCTGTTATCCAAGAGTGTGTGACAAAGATTATTGGGATGCTATTA  
 CYSARGTYRCYSALAPROPROGLYTYRALALEULEUARGCYSASNASPTHRASNTYRSE  
 GATGTAGATATTGTGCACCTCCAGGTTATGCTTGTGCTAGATGTAATGACACAAATATT  
 6800  
 GLYPHEMETPROASNLCYSERLYSVALVALVALSERSERCYSTHRARGHETMETGLUTHR  
 CAGGCTTTATGCCTAAGTCTAAGGTAGTGGTCTCTTCATGCACAAGGATGATGGAGA  
 6900  
 GLNTHRSERTHRTRPPHEARGPHEASNGLYTHRRARGALAGLUASNARGTHRTYRILETYR  
 CACAGACTTCTACTTGGTTTCGGTTAATGGAAGTACAGCAGAAAATAGAACCTATATT  
 TRPHISGLYARGASPNARGTHRILEILESERLEUASNLYSHISTYRASNLEUTHRMET  
 ACTGGCATGGTAGAGATAATAGGACTATAATTAGTCTAAATAAGCATTATAATCTAACAA  
 7000  
 LYSCYSARGARGPROGLYASNLYSTHRVALLEUPROVALTHRILEMETSERALALEUVAL  
 TGAATGTAGAAGACCAGGAAATAAGACAGTTTACCAGTCACCATTATGCTGCATTGG  
 PHEHISSEGLNPROVALASNGLUARGPROLYSGLNALATRPCYSARGPHEGLYGLYASN  
 TTTTCCACTCACAACCAGTCAATGAGAGGCCAAAGCAGGCATGGTGTAGCTTGGAGGAA  
 7100  
 TRPLYSGLUALALELYSGLUVALLYSGLNTHRILEVALLYSHISPROARGTYRTHRGLY  
 ATTGGAAGGAGGCAATAAAGAGGTGAAGCAGACCATTTCTCAAAATCCAGGTATCTG  
 7200  
 THRASNASNTHRASPLYSILEASNLEUTHRALAPROARGGLYGLYASPPROGLUVALTHR  
 GAACTAACAATACTGATAAAATCAATTTGACGGCTCTAGAGGAGGAGATCCGGAAGTTA

(fig.1B-suite 5)



PHE4ETTRPTHRASNCSARGGLYLUPHELEUTYRCYSLYSMETASNTRPPHELEUASN  
 CCTTCATGCTGGACAAATTGCAGAGGAGAGTTCTCTACTGTAAAATGAATTGGTTTCTAA  
 7300  
 TRPVALLUASPARGSERLEUTHRTHRGNLNLYSPROLYSGLUARGHISLYSARGASNTYR  
 ATTGGGTAGAAGATAGGAGTCTAACTACCCAGAAGCCAAGGAACGGCATAAAAGGAATT  
 VALPROCYSHISILEARGGLNILEILEASNTHRTRPHISLYSVALGLYLYSASNVALTYR  
 ACGTACCATGTCATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCAAAAATGTTT  
 7400  
 LEUPROPROARGGLUGLYASPLEUTHRCYSASNSETRHRVALTHRSELEULEALASN  
 ATTTGGCTCCAAGAGAGGGAGACCTCACGTCTAACTCCACAGTGACCAGTCTCATAGCAA  
 7500  
 ILEASNTRPTHRASPGLYASNGLNTHRSERILETHRMETSERALAGLUVALALAGLULEU  
 ACATAAAATGGACTGATGGAAACCAAACTAGTATCACCATGAGTGCAGAGGTGGCAGAAC  
 TYRARGLEUGLULEUGLYASPTYRLYSLEUVALGLUILETHRPROILEGLYLEUALAPRO  
 TGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAAATCACTCCAATTGGCTTGGCCC  
 7600  
 THRASNVALLYSARGTYRTHRTHRGLYGLYTHRSERARGASNLYSARGGLYVALPHEVAL  
 CCACAAATGTGAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAAGAGGGGTCTTTG  
 LEUGLYPHEUGLYPHELEUALATHRALAGLYSERALAMETGLYALALASERLEUTHR  
 TGCTAGCGTTCTTGGGTTTCTCGCAACGGCAGGTCTGCAATGGCGCGGGCGTCTGTA  
 7700  
 VALTHRALAGLNSERARGTHREULEUALAGLYILEVALGLNGLNGLNGLNGLNLEULEU  
 CCGTGACCGCTCAGTCCCGGACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGT  
 7800  
 ASPVALVALLYSARGGLNGLNGLULEULEUARGLEUTHRVALTRPGLYTHRLYSASNLEU  
 TGGACGTGCTCAAGAGACACAAGAATTGTTGCGACTGACCGTCTGGGCAACAAAGAACC  
 GLNTHRARGVALSERALAILEGLULYSTYRLEULYSASPGLNALAGLNLEUASNALATRP  
 TCCAGACTAGGGTCTCTGCCATCGACAAGTACTTAAAGGACAGGGCAGCTAAATGCTT  
 7900  
 GLYCYSALAPHEARGGLNVALCYSHISTHRTHRVALPROTRPPROASNALASERLEUTHR  
 GGGGATGTGGCTTTAGACAAGTCTGTCACACTACTGTACCATGGCCAAATGCAAGTCTAA  
 PROASPTRPASNASNLUTHRTRPGLNGLUTRPGLUARGLYSVALASPHELEUGLUALA  
 CACCAGATTGGAACAATGAGACTTGGCAAGAGTGGGAGCGGAAGGTTGACTTCTTGGAGG  
 8000  
 ASNILETHRALALEULEUGLUGLUALAGLNILEGLNGLNGLULYSASNMETTYRGLULEU  
 CAAATATAACGGCCCTCCTAGAAGAGGCACAAATTCAACAAGAGAAGAACATGTATGAAT  
 8100  
 GLNLYSLEUASNSETRPASPVALPHEGLYASNTRPPHEASPLEUTHRSERTRPILLYS  
 TACAAAAGTTGAATAGCTGGGATGTCTTGGCAATTGGTTTGACCTTACTTCTTGGATAA  
 TYRILEGLNTYRGLYILETYRILEILEVALGLYVALILELEULEUARGILEVALILETYR  
 AGTATATACAATATGGAATTTATATAATTGTAGGAGTAATACTGTTAAGAATAGTGATCT  
 8200  
 ILEVALGLNMETLEUALAARGLEUARGGLNGLYTYRARGPROVALPHESESERSETRPRO  
 ATATAGTACAAATGCTAGCTAGCTAAGACAGGGGTATAGGCCAGTGTCTCTTCCCCAC  
 TATZARGPROILEPROASNARGILEARGLEUCYSGLNPROLYSVALA  
 ART2VALASPPROTYPRTDTHRGLYSERGLYSERALAASNGLNARGARGGLN  
 SERTYRPHGLN\*\*\*THRMISTHRGLNGLNASPPROALALEUPROTHRLYSGLUGLYLYS  
 CCTCTTATTTCCAGTAGACCATACCCAACAGGATCCGGCTCTGCCAACCAAGAAGGCA  
 8300  
 LYSLSGLUTHRVALGLUALALAYALALATHRALAPROGLYLEUGLYARGTAT(f1n)  
 LYSARGARGAGTRPARGGLNARGTRPGLNGLNLEULEUALALEUALAASPARGILETYR  
 LYSGLYASPGLYGLYGLYSERGLYGLYASNSETRTRPPROTRPGLNILEGLUTYRILE  
 AAAAGGAGACGGTGGAGGCAGCGGTGGCAACAGCTCCTGGCCTTGGCAGATAGAATATA  
 8400

(fig.1B-suite 6)

SERPHEPROASPPROPROTHRASPTHRPROLEUASPLEUVALALEGLNGLNLEUGLNASH  
 HISPHELEUILEARGGLNLFUILEARGLEULEUTHRTPLPEUPHESERASNCYSARGTHR  
 TTCATTTCCTGATCCGCCAACTGATACGCCTCTTGACTTGGCTATTACCAACTGCAGAA  
 LEUVALILEGLUSERILEPROASPPROPROTHRASNILEPROGLUALALEUCYSASPLFU  
 LEULEUSERARGALATYRGLNILELEUGLNPROILEPHEGLNARGLEUSERALATHRTYR  
 CTTTCTATCGAGAGCATACCAGATCCTCCAACCAATATTCCAGAGGCTCTCTGGACCT  
 8500 F METGLYGLYALA  
 ARGARGILEARGGSRPROGLNALA ● ART2 (fin)  
 GLYGLUPHEGLYGLUVALLEUARGLEUGLULEUTHRTYRLEUGLNTRYRGLYTRPSERTYR  
 ACCGAGAATTCCGAGAAGTCTCAGGCTTGAACCTGACCTACCTACAATATGGGTGGAGCT  
 ILESERLYSLYSARGSERLYSPROPROGLUILECYASASPARGASPSERCYSGLYARGVAL  
 PHEGLNGLUALALGLNALAALAARGASPLEUARGGLNARGLEULEUARGALAAARGGLY  
 ATTTCCAAGAAGCGGTCCAAGCCGCCAGAGATCTGCCAGAGAGACTCTTGGGGCGCGTG  
 8600  
 GLYARGASNTYRGLYARGLEUPHELYSGLYVALGLUASPLYSERSEGLNLSERLEUGLY  
 GLULYSLEUTRPLGLUALALEUGLNARGGLYGLYARGTRPILELEUALALEPRDARGAR  
 GGGAGAAATTATGGGAGGCTCTTCAAAGGGGTGGAAGATGGATCCTCGCAATCCCTAGGA  
 8700  
 GLTLEUASPLYSGLYLEUSERSERLEUSERCYSGLUGLYGLNLYSTYRASNGLNGLYGLU  
 ILEARGGLNGLYLEUGLULEUTHRLEULEU ●  
 GGATTAGACAAGGGCTTGAGCTCACTCTCTTGTGAGGGCCAAAAATACAATCAGGGAGAA  
 TYRMETASNTHRPROTRPARGASNPRODALAGLUGLUARGLYSLYSLEUPROTYRARGLYS  
 YACATGAATACTCCATGGAGAAACCCAGCTGAAGAGAGGAAAAATTACCATACAGAAAA  
 8800  
 GLNASNILEASPAPILEASPLUGLUASPAASPLEUVALGLYILEPROVALGLUALA  
 CAAAATATAGATGATAGATGAGGAAGATGATGACTTGGTAGGGATACCAGTTGAGGCC  
 ARGVALPROLEUARGTHRMETSERTYRLYSLEUALALEASPMETSERHISPEILELYS  
 AGAGTTCCCTAAGAACAATGAGTTACAAATTGGCAATAGATATGTCTCATTTTATAAAA  
 8900  
 GLULYSGLYGLYLEUGLUGLYILETYRTRYRSEALAAARGARGHISARGILELEUASPILE  
 GAAAGGGGGGCTGGAAGGATTTATTACAGTGAAGAAGACATAGAACTTAGACATA  
 9000  
 TYRLEUGLULYSGLUGLUGLYILEILEPROASPTRPGLNILEHISSEGLYPROGLYILE  
 TACTTAGAAAAGGAAGGATCATACCAGATTGGCAGATACACTCCGAGCAGGAATT  
 ARGTYRLEULYSMETPHEGLYTRPLEUTRPLYSLEUITLEPROVALASNYALSERASPLU  
 AGATACCTAAAGATGTTTGGCTGGCTATGGAATTAATCCCTGTAATGTATCAGATGAG  
 9100  
 ALAGLNGLUASPLUGLUMISTYRLEUVALHISPRODALAGLNTHRSERGLNTRPASPASP  
 GCACAGGAGGATGAGGAGCATTATTTAGTGACCCAGCTCAAACCTCCAGTGGGATGAC  
 PROTRPGLYGLUVALLEUALATRPLYSPEASPPROTHRLEUALATYRTHRTYRGLUALA  
 CCTTGGGGAGAGGTTCTAGCATGGAAGTTTGATCCAACCTCTAGCCTACCTTATGAGGCA  
 9200  
 TYRILEARGTYRPROGLUGLUPHEGLYSERLYSSERGLYLEUSERGLULYSGLUYALLY  
 TATATTAGATACCCAGAAGAGTTTGAAGCAAGTCAGGCTGTCTAGAGAAAGAGGTTAAA  
 9300  
 ARGARGLEUALAALAARGGLYLEULEUGLUMETALAASPARGLYGLUTHRSER  
 AGAAGGCTAGCCGCAAGAGGCTTCTTGAAATGGCTGACAGGAAGGAACTAGCTGAGAC  
 AGCAGGGACTTTCCACAAGGGGATGTCTGAGGGAGGACTGGGAGGAGCGGTTGGGAA  
 9400  
 CACCCACTTTCTTGATGTATAAATATCACTGCAATTCGCTCTGTATTCACTCGCTCTGCG  
 GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCTCCAGCACTAGCAGGTAGAGCCTGGG  
 9500  
 TGTTCCTGCTAGACTCTCACCAGCACTTGGCCGGTCTGGCCAGAGTGGCTCCACGCTT  
 9600

(fig.1B-suite 7)



```

-----AGCTGAGACAGCAGGGACTTTCCACAAGGGGATGTCATG--GGGA
          9360      9370      9380      9390
          : : : : : : : : : : : : : : : : : : : : : :
GTAAC TAACAGAAACAGCTGAGACTGCAGGGACTTTCCAGAAGGGGCTGTAACCAAGGGA
          9370      9380      9390      9400      9410

          9400      9410      9420      9430      9440      9450
GGTACTGGGGAGGAGCGGTTGGGAACACCCACTTTCTTGATGTATAAATATCACTGCAT
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
GGGACATGGGAGGAGCTGCTGGGAACGCCCTCATATTCTCTGTATAAATATACCCGCTA
          9430      9440      9450      9460      9470

          9460      XX      10      20      30      40
TTCGCTCTGTA--TTCTGGAAGGGATTATTACAGTGCAAGAAGACATAGAATCTTAGAC
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
GCTTGCAATTGTAATTCTGGAAGGGATGTTTACAGTGAAAGAAGACATAAAATCTTAAAT
          9490      XX      10      20      30      40

          50      60      70      80      90
ATATACTTAGAAAAGGAAGAAGGCATCATACCAGATTGGCAGATACACTCCGGA---CCA
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
ATATACTTAGAAAAGGAAGAAGGGATAATTGCAGATTGGCAGAACTACACTCATGGGCCA
          50      60      70      80      90      100

          110      120      130      140      150
GGAATTAGATACCTAAAGATGTTTGGCTGGCTATGGAATTAATCCCTGTAAATGTATCA
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
GGAGTAAGATACCCAATGTTCTTTGGGTGGCTATGGAAGCTAGTACCAGTAGATGTCCCA
          110      120      130      140      150      160

          170      180      190      200      210
GATGAGGCACAGGAGGATGAGGAGCATTATTTAGTGCACCCAGCTCAAACCTTCCCAGTGG
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
CAAGAAGGGGAGGACACTGAGACTCACTGCTTAGTACATCCAGCACAAACAAGCAAGTTT
          170      180      190      200      210      220

          230      240      250      260      270
GATGACCCCTGGGGAGAGGTTCTAGCATGGAAGTTTGATCCAACCTCTAGCCTACACTTAT
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
GATGACCCGCATGGGGAGACACTAGTCTGGGAGTTTGATCCCTTGCTGGCTTATAGTTAC
          230      240      250      260      270      280

          290      300      310
GAGGCATATATTAGATACCCAGAAGAGTTTGGAAGCA
: : : : : : : : : :
GAGGCTTTTATTCGG
          290

```

(fig.1C-suite 1)

FIG. 2  
( HIV-2.P  
( versus  
( HIV-1.P

```

.....
HIV2----- 10 20 30 40 50 env4
          ----- MNNQLLIA ILLA-SACLV YCTAYVTVY GVPVWKEATI
HIV1----- HRVKEKYQHL WRNGWRWGTN LLGILNICS A TEKLWTVVY GVPVWKEATT
.....
HIV2----- 60 70 80 90 100 env5
          ----- PLFCATRRR -DT- WG TIQCLPDHDD YOEITL-HVI EAFDAVNNIV
HIV1----- TLFCASDAKA YDTEVHVWVA THACVPTDPN PQEVVLVNV I ENFNMMKNDM
.....
HIV2----- 110 120 130 140 150 env6
          ----- TEQAIEDVWH LFETSJHPCV KLTPLCVANKI CSSTESSIGN NTTSKSTSTT
HIV1----- VEQNHEDIIS LWDQSLKPCV KLTPLCVSLK CTDL---GN ATNTNSSNTN
.....
HIV2----- 160 170 180 190 200
          ----- --TTITPDQE QEISEDTPCA RADNCSGLGE EETINCQFNM TGLERDKKKQ
HIV1----- SSSGENHMEK GEIK----- NCSFNIS TSIRGKVQKE YAFFYKLBII
.....
HIV2----- 210 220 230 240 250 env7
          ----- Y--NET-WYS KVVCEINNST NQTQCYHNC NTSVITESCD KHYWDAIRER
HIV1----- PIDNDTTSYT ----- TSC NTSVITQACP KVSFEPIPIH
.....
HIV2----- 260 270 280 290 300 env8
          ----- YCAPPGYALL RC-NDI-WYS GFAPNCSKV ASTCTRMHET QTSTWF-GFN
HIV1----- YCAPAGFAIL KCNNKTFNGI GP---CTNVS TVQCTHGIRP VVSTQLLL-H
.....
HIV2----- 310 320 330 340 350
          ----- GTRAE---H RTYIYWBCRD H-RTH-SLN KYYNLSLHCK RPNKHTVKQI
HIV1----- GSLAEDEVVI RSNFT---D NAKTIIVQLN QSVE--INCT RPNNHTRRSI
.....
HIV2----- 360 370 380 390 400 env9
          ----- HLNS--GHVF ESHYQPIKRR PROANCWFKG -EUKDANQEV KETLAKHPRY
HIV1----- RIQRCPGRAF VTICKICH-- HRQANCNISR AKWHAAT---L KQIASKLREQ
.....

```

FIG. 2

```

                                410      420      env10      430      440      450
HIV2----- RGTNDTRNIS FAAPGKGS DP EVAYMWTNCR GEFLYCNMTW FLH--WI---
              * * * * *
HIV1----- FGNHKT--II FKQSS-GGDP EIVTHSFNCG GEFFYCNSTQ LFNSTWPNST
.....

                                460      470      env11      480      490      500
HIV2----- --EN KTHENYAPCH IKOIINTWHK VGRNVYAPPA EGELSCNSTV
              * * * * *
HIV1----- WTEGSSNNIE GSDTITLPCR IKQFINHWQE VGKAMYAPPI SGQIRCSSNI
.....

                                510      520      530      540      550
HIV2----- TSIIANIDWQ NNNQTNITFS AEVAELYRL- --ELGDYKLV EITPICFAPT
              * * * * *
HIV1----- TGLLLTRDGG NNNNGSEIFR PGCGDHRDNW RSELYKYKVV KIEPLGVAPT
.....

                                env3      560      570      580      590      600
HIV2----- KEKRYSSAHC RHTRGVFVLC --FLCFLATA GSAMGAAS-- LTVSAQSRTL
              * * * * *
HIV1----- KAKRR--VVQ. REKRAVCI-G ALFLGFLCAA CSTHGARSMT LTVQA--RQL
.....

                                610      620      630      640      env1      650
HIV2----- LAGIVQQQQQ LLDVVKRQQE LLRLTVWGTE NLQARVTAIE KYLODOARLN
              * * * * *
HIV1----- LSGIVQQQNN LLRAIEAQQE LLQLTVWGIK QLQARILAVE RYLKDQQLLG
.....

                                660      670      680      690      700
HIV2----- SHGCAFRQVC HTIVPW-- VNDSLAPDND NMTWQEWKQ VRYLEANISK
              * * * * *
HIV1----- IWGCSGKLIC TTAIVFWNASI SNNSELIQWN NMTWMEWDRE INNYTSLINS
.....

                                ↓      710      env2. 720      730      740      750
HIV2----- SLEQAQIQQE KNMYELQKLN SVDIFGNWFD LTSWVKYIQY GVLIIIVAVIA
              * * * * *
HIV1----- LIEESQNQQE XHEQELLELD KWASLWNWFB ITNWLWYIKI FIMIVGGLVG
.....

```

(fig.2 - suite 1)

```

              760      770      780      790      800
HIV2----- LRIVYVVQH LSLRKGYRP V-FSSPPGYI QQIHINKDRC QPANEEETED
          **** *      * * * * *      *      **      * **
HIV1----- LRIVFAVLSI VNRVRQCYSF LSFQT----- --HLPTPRG PDRPEGIEEE
.....

              810      820      830      840      850
HIV2----- GGSNGCDRYW PWPIAYIHFL IRQLIRLLT- ---LYSIC RDLLSRSLT
          **      **      *      * *      ****
HIV1----- GGERDRDRSI RLVNGSLA-L IWDDLRLSLCL FSYHRL--- RDLLLVTRI
.....

              860      870      880      890      900
HIV2----- LQLIYQNLRD WLRLRTA--F LQYGCEWIQE AFQ---AAA RATRETL---
          *      *      *      ***      * *      * *
HIV1----- VELLG--RRG WEALKYWNL LQYWSQELKN SAVSLNATA IAVAECTDRV
.....

              910      920      930      938
HIV2----- ---AGACRG LWRVLERIGR GILAVPRRIE QGAELALL
          **      ***** ** * **
HIV1----- IEVVQGACRA ----- -IRHPRRIE QGLERILL
.....

```

(fig. 2 - suite 2)

FIG. 3 ( ENV-mac  
( versus  
( ENV-ROD

```

      10      20      30      40      50
MGCLGNOLLIAIC--SKCLWIIICIQYTVFYGVPAWRNATIPLFCA TKNRDTWGTQCL
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
MM-----NQLLIAILLASACLVY-CTQYTVFYGVPTWKNATIPLFCA TRNRDTWGTIQCL
      10      20      30      40      50

      60      70      80      90     100     110
PDNDYSELALNYTESFDAWENTVTEQAI EDVWQLFETS IKPCVKLSPLCITHRCNKSET
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
PDNDYQEITLNVTEAFDAWNTVTEQAI EDVWHLFETS IKPCVKLTPLCVAMKCSSTES
      60      70      80      90     100     110

      120     130     140     150     160     170
DKHGLTKSS TTTASTTTTAKSVETRDIVNETS--PCVVHDNCTGLEQEP MISC KFNH
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
STGNNTSKST--STTTTTP-----T-DQEQEISEDTPCARADNC SGLGEEET INCOFNM
      120     130     140     150     160

      180     190     200     210     220     230
TGLKRDKKKEYNETHWYSADLVCEQGNSTGNESRCYMNHCNTSVIQECCDKDYWD A IRCRY
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
TGLERDKKKQYNETHWYSKDVVCETNNST-NQTCYMNHCNTSVITESCKHYWD A I RFRY
      170     180     190     200     210     220

      240     250     260     270     280     290
CAPPGYALLRCNDTNYSGFMPNCSKVYVSSCTRM METQTSTWFRFNGTRAENRTYIYWHG
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
CAPPGYALLRCNDTNYSGFAPNCSKVVA STCTRM METQTSTWFGFNGTRAENRTYIYWHG
      230     240     250     260     270     280

      300     310     320     330     340     350
RDNRTIISLNKHYNLTHKCR RPGNKTVLPVTIMSALVFHS--OPVNERPKOAWCRFGGNW
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
RDNRTIISLNKY NLSLHCKRPGNKTYKQIMLSGHV FHSHYOPINKRPROAWCWFKGKW
      290     300     310     320     330     340

      360     370     380     390     400
KEAIKEYKQTI V KHPRYTG TNNTDKINLTAPRGG-DPEVTFMWTNCRGEFLYCKMWNFLN
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
KDAHQEVKETLAKHPRYRG TNDTRNISFAAPGKGS DPEVAYMWTNCRGEFLYCNMTWFLN
      350     360     370     380     390     400

```

FIG. 3



```

      420      430      440      450      460
WVEDRGLTQPKERHKRNYVPCHIRQIINTWHKVGKNVYLPPREGDLTCNSTVTSLIAN
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
WIEN-----KT-H-RNYAPCHIKQIINTWHKVGKNVYLPPREGELSCNSTVTSLIAN
      410      420      430      440      450

      480      490      500      510      520
INWTDGNQTSITMSAEVAELYRLELGDYKLVEITPIGLAPTNVKRYTTG-GTSRNKRGVF
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
IDWONNNQTNITFSAEVAELYRLELGDYKLVEITPIGFAPTKEKRYSSAHG--RHTRGVF
      460      470      480      490      500      510

      540      550      560      570      580
VLGFLGFLATAGSAMGAASLTVTAQSRLLAGIVOOOQOLLDVYKROQELLRLTVWGTKN
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
VLGFLGFLATAGSAMGAASLTVSAQSRLLAGIVOOOQOLLDVYKROQELLRLTVWGTKN
      520      530      540      550      560      570

      600      610      620      630      640
LOTRVSAIEKYLKDOAOLNAWGCAFRQYCHTTVPWPNASLTPDWNNEHWQEWKRVDFLE
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
LQARVTAIEKYLODQARLNSWGCAFRQYCHTTVPWVNDLAPDWDNMHWQEWKQVRYLE
      580      590      600      610      620      630

      660      670      680      690      700
ANITALLEEAQIQEKNMYELOKLNWDVFGNWFDLTSWKYIQYGIYIIVGVILLRIVI
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
ANISKSLEQAQIQEKNMYELOKLNWDVFGNWFDLTSWKYIQYGLIIVAVIALRIVI
      640      650      660      670      680      690

      720      730      740      750      760
YIVQMLARLRGYRPVFSSPPSYFQ*THTOQDPALPTKEGKKGDGGGSGGNSSWPHQIEY
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
YVVQMLSRLRGYRPVFSSPPGYIQQIHMKDRGOPANEETEDGGSNGGDRYWPWPYAY
      700      710      720      730      740      750

      780      790      800      810      820
IHFLIRQLIRLLTWLFSNCRTLLSRAYQILQPIFORLSATYGEFGEVLRLELTLYQYGWS
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
IHFLIRQLIRLLTRLYSICRDLLSRSFLLTOLYONLRDW-----LRLRTAFQYGC
      760      770      780      790      800

      840      850      860      870      880
YFOEAVQAA-RDLRORLLRA-RGEKLWEALQGGRMILAIIPRRIROGLELTLL
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
WIQEAFQAAARATRETLACRG--LWRVLERIGRILAVPRRIROGAETALL
      810      820      830      840      850

```

(fig. 3-suite 1)

FIG. 4

EP 0 750 041 A2

```

420      430      440      450      460      470
EGHSAROCRAPRROGCHKCGKMDHYMAKCPNROAGFLGLGPHGKKPRNFPMAQVMQGLTP
::::::::::::::::::::: : : : : ::::::::::::::::::::::: : : : :
EGHSAROCRAPRROGCHKCGKPGHIMTNCPDROAGFLGLGPHGKKPRNFPVAQVPOGLTP
      400      410      420      430      440      450

480      490      500      510
TAPPEEPAYDLLKNYMHGKQQRERSGKPYKEVTEDLLHL-----NS
:::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TAPPYDPAYDLLEKYMOQGKRQREORERPYKEVTEDLLHLEQGETPYREPPTEDLLHLNS
      460      470      480      490      500      510
```

(fig.4 - suite 1)

FIG. 5

410	420	430	440	450	460
DRVVLQKELLNSIGFSSPEEKFOKDPFQWMGYELWPTKWKLOKIELPORETWTVNDIO					
410	420	430	440	450	460
DRVVLQKELLNGLGFSTPDEKFOKDPFQWMGYELWPTKWKLOKIQLOPOKEIHTVNDIO					
390	400	410	420	430	440
470	480	490	500	510	520
KLVCYLNWAAQIYPGIKTKHLCLIRGKNTLTEEVOVTEMAEAEYEENKIIILSDEQEGCY					
470	480	490	500	510	520
KLVCYLNWAAQIYPGIKTKHLCLIRGKNTLTEEVOVTEMAEAEYEENRIILSDEQEGHY					
450	460	470	480	490	500
530	540	550	560	570	580
YQESKPLEATYIKSDQNQWSYKIHQEDKILKVYKFAKIKNTHNGVRLLAHVIOKIGKEA					
530	540	550	560	570	580
YQEEKELEATYOKDQENOWTYKIHQEEKILKVYKFAKIKNTHNGIRLLAQVYKIGKEA					
510	520	530	540	550	560
590	600	610	620	630	640
IYIWGCVKPKFHLVPEKDYQWMTDYWQVTHIPEWDFISTPPLVRLVFNLYVKDPIEGEET					
590	600	610	620	630	640
LYIWGRIPKFLPLVERIWEQWMDNYWCYTHIPEWDFYSTPPLVRLAFNLVGDPIPCAET					
570	580	590	600	610	620
650	660	670	680	690	700
YYVDGSCSKQSKGKAGYITDRGKDKVYKLEQTTNOOAELEAFMALTDSPKANIIIVDS					
650	660	670	680	690	700
FYTDGSCNROSKEGKAGYITDRGKDKVYKLEQTTNOOAELEAFMALTDSPKVNIIIVDS					
630	640	650	660	670	680
710	720	730	740	750	760
QYVMGIIIGCPTESERLYNQIIEEMIKKTEIYVAVVPAHKGIGGNOEIDHLYSOGIROV					
710	720	730	740	750	760
QYVMGISASOPTESERLYNQIIEEMIKKTEIYVAVVPAHKGIGGNOEIDHLYSOGIROV					
690	700	710	720	730	740
770	780	790	800	810	820
LFLEKIEPAQEHSHYSNINELVFKGLPRLVAKQIVDTCDKCHQKGEAIGHQVNSDLG					
770	780	790	800	810	820
LFLEKIEPAQEHSHYSNINELVFKGLPRLVAKQIVDTCDKCHQKGEAIGHQVNAELG					
750	760	770	780	790	800
830	840	850	860	870	880
THOMDCTHLECKIIVAVVHVASGFIEAEVPOETGROTALFLLKLASRWPITHLHTONGA					
830	840	850	860	870	880
THOMDCTHLECKIIVAVVHVASGFIEAEVPOESGROTALFLLKLASRWPITHLHTONGA					
810	820	830	840	850	860
890	900	910	920	930	940
NFASQEVKMYAHWAGIEHTFGVPYNPQSQGVVEAMNHHLKNQIDRIREQANSYETIVLMA					
890	900	910	920	930	940
NFTSQEVKMYAHWAGIEHTFGVPYNPQSQGVVEAMNHHLKNQISRIREQANTJETIVLMA					
870	880	890	900	910	920
950	960	970	980	990	
VHCMNFKRRGGIGDMTPAERLINMIITEQEIQFOQSKNSKFKNFRVYVYFREGRDQLWKGP					
950	960	970	980	990	
IHCMNFKRRGGIGDMTPSERLINMIITEQEIQFLOAKNSKLKDFRVYFREGRDQLWKGP					
930	940	950	960	970	980
1010	1020	1030	-1040	1050	
ELLWKGEVAVILKVGTDIKYVPRRKAKIIOYGGGKEMOSSSHHEDTGEAREYA					

**EP 0 750 041 A2**

```

.....:.....:.....:.....:.....:.....:.....:.....:.....:.....:
ELLWKGE GAVLVKVGTDIKIIPRRKAKIIRDYGGROEMDSGSHLEGAREDEGMA
      990          1000          1010          1020          1030

```

(fig. 5-suite 2)

**FIG. 6**

( R.mac  
FIG. 7 ( versus  
 ( R.ROD

	10	20	30	40	50
ME---	ERPPENEGPQREPHDEWVVEVLKELKEEALKHFDPRLLTALGNHIYNRHGDTLE				
:	:	:	:	:	:
MAEAPTELPPVDGTPLEPGDEWIIIEILREIKEEALKHFDPRLLIALGKYIYTRHGDTLE					
	10	20	30	40	50
60	70	80	90	100	
GAGELIRILORALFIHFRSGCSHSRIGOPGGGNPLSTIPPSRSM					
:	:	:	:	:	:
GARELIKVLORALFTHFRAGCGHSRIGOTRGCNPLSAIPTPRNMQ					
	70	80	90	100	

FIG. 7



55

FIG. 9 ( F.mac  
( versus  
( F.ROD

```

      10      20      30      40      50
MGGAI SKKR SKPPEICD-RDSCGRVGRNYGR LFK-GVEDGSSQSLGGLOKGLSSLSCEGO
::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
MGASGSKKHSRPPRGLQERLLRARAGACGGYWNESGGEYSRFQE--GSDREOKSPSCEGR
      10      20      30      40      50

      60      70      80      90      100     110
KYNQGEYMNTPHRNPAEERKKLPYRKQNIIDIDEEDDLVGIPYEARVPLRTHSYKLAID
::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
QYQGGDFMNTPHKDPAAEREKNLYRQQNMDDVSDDDDOVRVSVTPKVPLRPMTHRLAID
      60      70      80      90      100     110

      120     130     140     150     160     170
MSHFIKEKGGLEGIIYSARRHRILD IYLEKEEG IIPDWOI--HSGPGIRY LKMFGWLWKL
::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
MSHLIKTRGGLEGMFYSERRHKILNIYLEKEEG IADWONYTH-GPGVRYPMHFGWLWKL
      120     130     140     150     160     170

      180     190     200     210     220     230
IPVNVSDAEQDEEHYLVHPAOTSQWDDPNGEVLAWKFDPTLAYTYEAYIRYPEEFGSKS
::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
VPVDYPQEGEDTETHCLVHPAOTSKFDDPNGETLVWEFDPLLAYSYEAFIRYPEEFGHKS
      180     190     200     210     220     230

      240     250     260
GLSEKEVKRRLLAARGLLEMADRKETS
::  ::  ::  ::  ::  ::
GLPEEENKARLKARGIPFS
      240     250

```

FIG. 9

FIG.10 ( TAT.mac  
( versus  
( TAT.ROD

```

      10      20      30      40      50
METPLREQENSLESSNERSSYISEAAAAIPESANLGEEILSQLYRPLEACYNTCYCKKCC
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
METPLKAPESSLKSCNEPFSRTSEQDVATQELARQGEELSQLYRPLECNNSCYCKRCC
      10      20      30      40      50

      70      80      90      100      110
YHCQFCFLKKGLGISYEKSHRRRRTPKKAKANTSSASNERP---IPNRIRLCOPKKAKKE
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
YHCQMCFLNKGLGICYERKGRRRRTPKKTKTHPSPT----PDKSISTRTGDSQPTKKQKK
      70      80      90      100      110

      120      130
TVEAAVATAPGLGR
: : : : : : : :
TVEATVETDTGPR
      120      130

```

FIG. 10

FIG. 11 ( ART.mac  
( versus  
( ART.ROD

```

      10      20      30      40      50
MRSHTGEEELRRRLRLIMLLHQTSKYGLSNKSAAYRMLLVDPYPTGSGSANORRQKRRRW
:   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
MNERADEEGLORKLRLIRLLHQTN-----PYPGPGTASQRRNRRRRW
      10      20      30      40

      70      80      90      100     110
RORWOOLLALADRIYSFPDPPTDTPDLAIQQLONLAIESIPDPPTNIPALCDLRRIR
:   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KORWROILALADSIYTFPDPADSPDQTIOHLQGLTIQELPDPPTHLPESQRLAET
      50      60      70      80      90      100

```

SPQA

FIG. 11